The contribution of migrants to the exports of Spanish provinces: Implications from gravity modelling with panel data

The contribution of migrants to the exports of Spanish provinces:

Implications from gravity modelling with panel data

Anna D'Ambrosio*

June 27, 2014

Abstract

A sizeable academic literature has documented a positive effect of immigration in promoting trade and investment between the host and the origin country. Using province-level international exports data from Spain, I contribute to the migration-trade link literature in three ways. First, I include a focus on emigration along with immigration. Second, I integrate the latest methodological developments of the research on gravity models in the analysis of the migration-trade link to address estimation issues and potential functional form mis-specification in the previous literature. Third, implications from the use of sub-national units for theory-consistent modelling are discussed and made explicit. The results of the analysis lead to identify the Gamma PML as the consistent and efficient estimator for the data at stake and to discard the Poisson PML in spite of its growing popularity as a "workhorse" in gravity models. The estimates robustly confirm a positive and significant effect of province-level immigration stocks on the exports of Spanish provinces and highlight important sub-national variation in the effects; the dynamics of the effects of emigration on trade, instead, appear to be determined at the national level.

Jel classification: F10, F14, F22, C52

Keywords: Gravity model, immigration, emigration, network effect, subnational units, gamma regression

1 Introduction

The impact of migration flows on the host economies is a heatedly debated topic in politics and social sciences. Most visible and documented at the level of employment, migration

*PhD Candidate in Innovation and Local Development, University of Trento

also importantly impacts at the level of the internationalization of the host economies. A sizeable academic literature since Gould [1994] and Rauch and Trinidade [2002] has documented a positive effect of immigration in promoting trade and investment between the host and the origin country.

The analysis of the migration-trade link employs gravity models of international trade where the stocks of immigrants, as a proxy for the size of the immigrant networks, are interpreted as factors that reduce the informal barriers to trade between the origin and the destination country. The developments in the theoretical foundations of the gravity model have co-determined an evolution in the estimation methods employed for the analysis of the migration-trade link; however, in spite of the strong case for the use of the Poisson estimator in gravity models since the article by Santos-Silva and Tenreyro [2006], this estimator has not found much prominence in the migration-trade link literature. The reason is probably that, due to the specific problems at stake, recent papers in this strand of the literature have been more interested in panel data than the more standard gravity literature, which has focussed very much on determinants of trade that show less marked time variation such as trade agreements, currency unions, borders and distance; theory consistent estimation by panel data involves the use of three sets of dummies (time-varying importer effects; time-varying exporter effects; and dyadic time-invariant effects), which increases very much the computational difficulty of estimating gravity models by Poisson[1]. Thus, the first contribution of this paper to the literature is methodological: drawing on the options proposed by Head and Mayer [2014] as a guidance for "exploring the robustness" of the OLS, Poisson and Gamma estimates, I explore the estimation issues affecting each estimator, such as heterogeneity, incidental parameters problem and heteroskedasticity, as well as the potential functional form mis-specification in the migration variables. No single estimator emerges as free from estimation issues. Based on the application of the tests proposed in Manning and Mullahy [2001] and Santos-Silva and Tenreyro [2006], the OLS and Gamma estimators appear more appropriate for the specific data at stake.

The second contribution relates to the specific migration history of Spain: building on the intuition by Murat and Pistoiesi [2009] that the arguments on the migration-trade link could be applied to emigration as well as to immigration, I include the analysis of emigration within theory-consistent modelling of the migration-trade link.

A third contribution is associated with the focus on subnational units of analysis, which is not novel *per se* but rather in the way its implications for theory-consistent modelling are discussed and made explicit.

The paper is organized as follows. In section 2 I present the theoretical framework on gravity models and on the migration-trade link, I run a survey of the literature of closest

[1] In practice, obtaining Poisson estimates in specifications including a large number of fixed effects may be impossible, as noted among other by Bratti et al. [2012], due to serious convergence problems.

relevance to this paper and I outline a few hypotheses with respect to the effects of immigration and emigration on trade. In section 3 I describe the dataset. In section 4 I define the econometric strategy applied to address the hypotheses. In section 5 I present the main results and open issues arising from the econometric analysis. Section 6 concludes.

2 Theoretical framework

The literature on the migration-trade link can be seen as a subset of a wider international trade literature applying “augmented” gravity models. The literature on gravity models has seen a considerable development in the last two decades, especially as far as the theoretical foundations of the model are concerned [Head and Mayer, 2014]. The turning point in the theoretical micro-foundations of the gravity model is the theoretical derivation of the ”multilateral resistance term” by Anderson and van Wincoop [2003], which led to obsolescence of previous works applying the gravity model omitting this term. On this ground, Baldwin and Taglioni [2007] have established three common mistakes of the literature on gravity, what they term the ”gold”, ”silver” and ”bronze-medal mistakes”. Because the literature investigating the effects of migration on trade has closely followed the developments in the gravity theory, it is possible to run a critical review of this literature in the light of the developments in gravity theory. Before starting, however, I will give a brief introduction to the gravity model.

2.1 The Gravity Model

Introductory paragraphs on gravity models often start from the statement that the gravity model is one of the few examples of law-like behaviours in social sciences: bilateral trade is proportional to the economic ”masses” of the trading countries and inversely proportional to their bilateral trade barriers, primarily distance. Figure 1 shows an instance of this: gravity applies to the trade of the province of Madrid with EU countries in 2008.

In Fig. 1, the log of the exports of the province of Madrid to EU countries are plotted against the GDPs of these countries divided by the bilateral distance. The idea is that the EU is a sufficiently homogeneous subset of countries from the institutional point of view to reduce the bias from omitted variables. The slope of the regression line, which results from the simple regression of the log of the exports on the log of the ratio of GDP to distance, is 1.00051, and most observations lie reasonably close to it. The proportionality of the GDPs and the inverse proportionality of distance to exports confirm the analogy with the physics law of gravitation with the exception, which is common in the trade literature, that distance is not squared. The gravity equation directly deriving from the physics analogy,

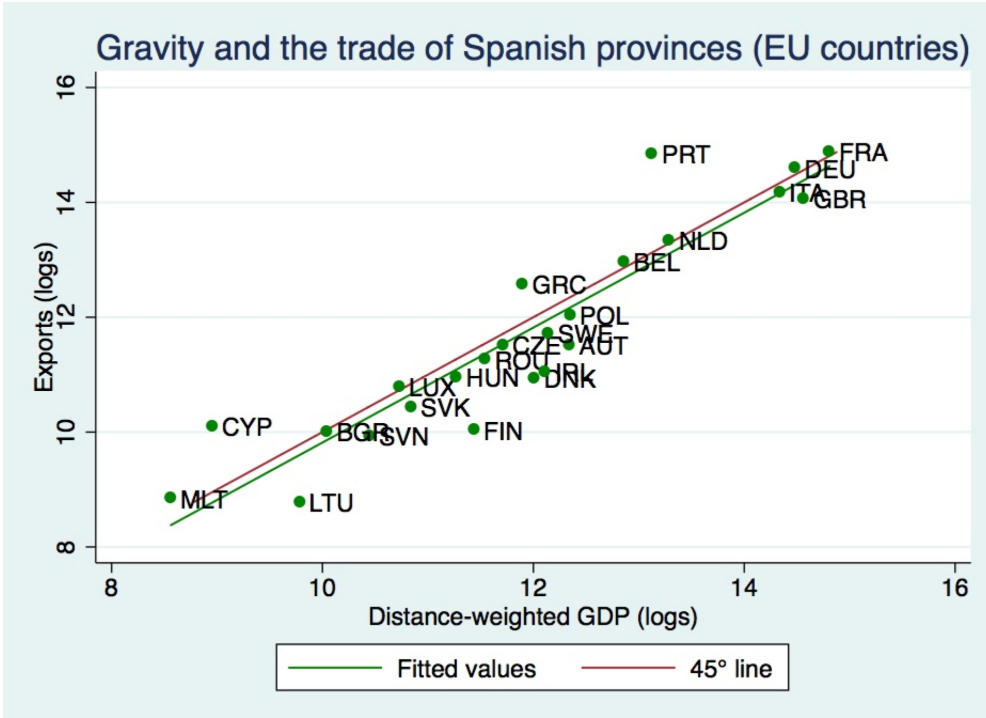


Figure 1: Gravity and the trade of the province of Madrid (EU countries)

empirically driven, and not theoretically micro-founded was first noted by Tinbergen [1962]; in their review, Head and Mayer [2014] call it the "naïve" gravity equation:

$$X_{ni} = GY_n^a Y_i^b \phi_{ni} \tag{1}$$

In Head and Mayer's notation, X_{ni} is the volume of trade between country n (importing country) and country i (exporting country); Y_i represents the "mass" of production of exporting country, Y_n represents the "mass" of expenditures of the importing country, G is a constant, and a and b are parameters to be estimated which do not necessarily equal 1 (even if empirically they are often close to this value). The term ϕ_{ni} represents the transaction costs of bilateral trade; it includes natural trade barriers such as distance but also other barriers such as tariffs, as well as their respective elasticities.

The theoretical foundations given to the gravity model in Anderson and van Wincoop [2003], building on Anderson [1979] made it clear that the naïve version of the equation was lacking a "multilateral resistance term". While the Anderson-Van Wincoop model is, by now, just one of the many compatible with what Head and Mayer [2014] call the "structural form" of the gravity equation, their derivation is useful to understand the meaning of the multilateral resistance terms. The Anderson-Van Wincoop model is derived in a general-equilibrium framework, where goods are differentiated by place of origin and each country produces only one good; preferences are approximated by a CES utility function, which consumers maximize under a standard budget constraint. Within the bundle of each consumer there are foreign-country products, whose price varies by country because of trade costs. The nominal demand for region i goods by region n consumers thus depends among other on trade costs between the two countries and on the price index in n. Assuming market clearance (and solving for the scaled prices), Anderson and Van Wincoop are able to derive an imports demand function that is analogous to what Head and Mayer [2014] label the "structural form" of the gravity model (see equation 4):

$$X_{ni} = \frac{y_n y_i}{y_w} \left(\frac{t_{ni}}{\Pi_i P_n} \right)^{1-\sigma} \tag{2}$$

In this equation, the income terms are expressed as income shares of world income $\frac{y_n y_i}{y_w}$ rather than as country income, P_n is the consumer price index of importing country n and Π_i is an exporter-specific term that is a function of all bilateral trade barriers and world income shares of its trade partners, t_{ni} are transport costs and σ is the elasticity of substitution between goods produced in different countries. Assuming symmetric trade barriers, both Π_i and P_n become price indices terms expressed as functions of all bilateral trade barriers and world income shares of each partner country. The gravity equation expressed in this way is the following:

$$X_{ni} = \frac{y_n y_i}{y_w} \left(\frac{t_{ni}}{P_i P_n} \right)^{1-\sigma} \tag{3}$$

The term $\left(\frac{1}{P_i P_n} \right)^{1-\sigma}$, which constitutes the main difference from the "naïve" version of the gravity equation, is called the "multilateral resistance term". The intuition is that bilateral trade depends on economic masses and relative trade barriers $\left(\frac{t_{ni}}{P_i P_n} \right)^{1-\sigma}$; thus, changes in bilateral trade costs influence trade between two countries to the extent that the change in bilateral costs is greater than the change in the average trade barriers of each country towards all its partner countries (Anderson and van Wincoop [2003]: 176). In this framework, assuming for example a uniform rise in trade barriers, two smaller countries which are very open to trade will face a larger effect on their multilateral resistance, and a smaller effect on their bilateral trade, than two larger countries: barriers will rise for all trade transactions of the small countries with the rest of the world, and because these make up an important part of the price indexes of the small countries, the multilateral resistance will increase. Because of the change in the multilateral resistance term, the bilateral change in trade barriers will be relatively unimportant and thus bilateral trade will not be much affected. On the contrary, larger countries for which the internal market is more important will see a relatively weaker effect of the increase in trade barriers on the internal price levels and thus on the multilateral resistance terms. Therefore, the bilateral increase in trade barriers with a given partner will have a bigger effect on bilateral trade because of the relatively smaller change occurred in the multilateral resistance term.

While the authors express the multilateral resistance through price indices in their model, they warn that the multilateral resistance terms should be considered in a wider sense to include non-pecuniary costs of trade and non-traded goods, which all contribute to explaining price differentials across countries and regions. This makes the multilateral resistance term unobservable, while part of it can be expressed by observables such as distance, income shares, tariffs, etc. For this reason, since Anderson and Van Wincoop, analysts no longer include GDP deflators into the gravity equation and rather resort to fixed effects.

A more general way of expressing Anderson-Van Wincoop's equation, abstracting from the general-equilibrium framework and from the assumption of symmetry in trade costs, is referred to by Head and Mayer [2014] as the "structural" gravity equation:

$$X_{ni} = \frac{Y_i}{\Omega_i}\frac{X_n}{\Phi_n}\phi_{ni} \tag{4}$$

Where the trade cost term ϕ_{ni} includes bilateral costs as well as their elasticities of substitution, and the multilateral resistance term corresponds to $(\Omega_i\Phi_n)^{-1}$. In turn,

$$\Omega_i = \Sigma_l \frac{\phi_{li}X_l}{\Phi_l} \text{ and } \Phi_n = \Sigma_l \frac{\phi_{nl}Y_l}{\Omega_l}. \tag{5}$$

It has been shown that the factors composing the multilateral resistance terms can be interpreted as, respectively, the average market access available to the exporting country (Ω_i) and the degree of competition in the importing country (Φ_n). More precisely, Ω_i represents the "expenditure-weighted average of relative access" and Φ_n the "accessibility-weighted sum of exporters' capabilities" ([Head and Mayer, 2014]: 9-10).

Thus, the total exporting capacity S_i of country i depends on its total supply Y_i and on the average market access available to it, Ω_i, which is a function of all other bilateral cost terms, the expenditure in all other countries, and the degree of competition in all other countries. The total importing capacity M_n of country n depends on its expenditures X_n and on the degree of competition in its market Φ_n. The actual trade between i and n depends ultimately on the two "monadic" terms S_i and M_n and on bilateral trade barriers ϕ_{ni}.

This is the logic underlying the third and last definition of the gravity equation, that of "general gravity":

$$X_{ni} = GS_iM_n\phi_{ni} \quad (6)$$

In this definition, the expenditure and multilateral resistance terms of each trading partner are captured in a unique ”monadic” term expressing the capacity of importing and of exporting of the two partners. This way of expressing the gravity equation makes it clearer why the recent literature captures both country expenditures (or supply) and multilateral resistance terms with fixed effects.

A variety of demand-side and supply-side models of international trade, including among other the Dixit-Stiglitz-Krugman monopolistic competition assumptions, Eaton and Kortum [2002], Helpman et al. [2008] and Chaney [2008], are compatible with the “structural” form of the gravity equation, and a number of other studies are compatible with its ”general” form. For a detailed review, see Head and Mayer [2014][2].

Baldwin and Taglioni [2007] notice that, with panel and time-series data, the multilateral resistance term depends on time-varying bilateral trade costs and on time-varying economic masses. Thus, it introduces an “un-constant” in the model, whose time variation should be accounted for in the empirical estimation. This bears important implications for empirical estimation: with cross-section data, a (usually) straightforward way to account for the multilateral resistance term is the inclusion of importer and importer fixed effects Anderson and van Wincoop [2003], Feenstra [2004], Redding and Venables [2004], Hummels [1999]. In a panel context, Baldwin and Taglioni [2007] argue that the time variation in the multilateral resistance terms should be coupled with importer-time and exporter-time effects, while the correlation between the unobservable component of the bilateral trade determinants and the included trade determinants should be accounted for by time-invariante pair effects. Omitting these terms introduces a substantial bias in the estimation and is referred to as the “gold-medal mistake” of gravity literature by Baldwin and Taglioni [2007][3].

In their Handbook chapter, Head and Mayer [2014], devote an entire section to the theory-consistent estimation of gravity equations and apply different methods in a series of Mon-

[2]Head and Mayer [2014] notice that all models compatible with the structural gravity form share two assumptions: one is market clearing on the export side; the second relates to the accounting identity by which total imports of country n from country i must equal the share allocated to country i of total expenditures in n: $X_{ni} = \pi_{ni}X_n$. The critical assumption is the multiplicative separability of π_{ni} as the product of three terms: one capturing bilateral trade costs, one capturing the degree of competition in the importing market (or multilateral resistance term for the exporter) and one capturing exporters’ capabilities (such as total production).

[3]The silver-medal mistake, often found in papers that use the average of bilateral flows as the dependent variable, relates to taking the log of the average of uni-directional flows rather than the average of the logs; the bronze-medal mistake relates to the inappropriate deflation of nominal trade values by the country’s aggregate price index: since there are global trends in inflation rates, inclusion of this term is likely to create biases via spurious correlations.

teCarlo simulation to identify the consistent one. Fixed effect estimation is the preferred solution "so long as there are no econometric issues"[4].

2.2 The mechanism: immigrants' contribution to trade

Within the "structural" gravity equation, migration stocks can be seen as one of the factors that reduce the bilateral costs of trade. Co-ethnic networks can be seen as a factor that reduces transaction costs related to information asymmetries and contract enforcement costs in international trade [Rauch and Trinidade, 2002]. Since the seminal work by Gould [1994], the literature on the US, UK and Canada has established a positive contribution of immigrants to trade and investments towards their origin countries (for example Dunlevy and Hutchinson, 1999, Wagner et al., 2002, Herander and Saavedra, 2005, Girma and Yu, 2002).

The reduction in information costs in bilateral trade attributable to immigrants and immigrant networks is generally referred to as the "information effect": it relies on the "deep" knowledge [Rauch, 2001] that immigrants have of their home country institutions, business culture and markets. This allows them play the role of intermediaries, provide market information, identify suitable suppliers or clients. The "enforcement effect" in international transactions relates to the contribution of community reputation mechanisms to ensure the enforcement of contracts [Rauch and Trinidade, 2002, Gould, 1994]. These mechanisms assume the activation of in-group social capital as a resource for business, in the sense of the "bounded solidarity" and "enforceable trust" described by Portes and Sensenbrenner [1993], that is generated by the relations of the community of co-ethnics with the host society. These lead to members' preferential access to in-group resources, to support by other members of the community, and to increased predictability in transactions due to shared norms.

The trade cost term ϕ_{ni} in equations 4 and 6 can also be broadly construed to include the

[4]The other estimation methods they consider are: *(i)* Estimation of the multilateral resistance terms through "structurally iterated least squares", a development of the non-linear least squares method used in Anderson and van Wincoop [2003]; *(ii)* Ratio-type estimation, where the monadic terms are suppressed algebraically thanks to the computation of export ratios of actual trade with either trade to self or with reference partners. The tetrad" method proposed by Head et al. [2010] computes a "ratio of ratios". It first divides the observed exports from country i to country n by the exports of exporter i to a reference importer k; then it divides the exports to n from a reference exporter ℓ by the exports from ℓ to k. Finally, the ratio of these two ratios must be computes. This way the monadic terms are eliminated; the trade cost terms must be similarly "tetraded", and the standard errors must be clustered at the importer, exporter and pair level. This methodology addresses the computational difficulty involved in fixed effects estimation with dummy variables but may be sensitive to the selection of the reference country; *(iii)* Double-demeaning of the variables, i.e. demeaning for the importer and exporter dimension, or combinations of de-meaning and dummy variables, and "Bonus vetus OLS", proposed by Baier and Bergstrand [2009], where first-order Taylor expansions of multilateral resistance terms are introduced in the specification.

measure of trust proposed by Guiso et al. [2009], which they find to affect bilateral trade in a significant way. Whether more sizeable immigration stocks affect the trust of the host country towards their origin country in a positive or a negative way cannot be determined a priori: Gould [1994], in his seminal work about the migration-trade link, argued that immigrants contribute to the overall diffusion in the host country, both to natives and non-natives, of information and knowledge about the institutions, business and language of the origin countries, and this can be interpreted as a positive contribution to bilateral trust. Hovewer, the social tensions that are in some cases associated with sizeable immigration stocks can have the opposite effect. In addition, levels of trust may be substantially different across sub-national units according to the size of the local immigrant communities.

Information and enforcement costs increase in the (cultural/institutional) distance between the trading countries[Boisso and Ferrantino, 1997, Dunlevy, 2006]. It is also possible that, in global trade transactions, the effect of somatic and genetic distance on bilateral trust and thus on trade is ceteris paribus stronger that what found by Guiso et al. [2009] for European countries. For these reasons, the effect of migration stocks on trade costs, and ultimately on trade volumes, is likely to be stronger for more culturally and institutionally distant trading pairs, as found in empirical studies by Girma and Yu [2002], Dunlevy [2006], Murat and Pistoiesi [2009]. In this sense, according to Tadesse and White [2008], migrants effectively offset the negative effects of cultural distance on exports[5].

In a study on the migration-trade link based on Chaney [2008]'s model, Peri and Requena-Silvente [2010] distinguish between fixed and variable costs of trade, arguing that immigrants are likely to affect both types of costs. Finding stronger evidence of the immigrants' effect on the extensive than on the intensive margin, they argue that immigrants contribute more to reducing the fixed costs of entering a foreign market than the variable costs of trade.

Besides trade costs, migration influences trade also trough the preferences channel (i.e., with reference to equation 4, by directly increasing X_n rather than reducing ϕ_{ni}). Since Gould [1994], this effect is referred to as a "preference effect", whereby immigrants affect imports of their host countries: immigrants are more likely than natives to consume their home country products and thus directly raise the demand for home country products

[5] Trade costs, in particular information costs, also have a diversified impact on different types of sectors and goods, and immigrants have a stronger effect on the trade of goods that have more informational content. Since Rauch and Trinidade [2002], many authors have found evidence of a stronger effect of immigrant stocks on the trade of more differentiated goods. The underlying mechanism is that, while for internationally homogeneous and reference-priced goods, price is conveying most of the relevant information for trade, exchanges of differentiated goods require specific efforts to identify and promote the specific features of the good and to match "buyers and sellers in the characteristics space" Rauch and Trinidade [2002]:117). The data used in this paper are not disaggregated by type of good and do not allow testing this relationship. It is however an ubiquitous and established finding in the literature, which has been confirmed by Peri and Requena-Silvente [2010] in the case of Spain on a slightly earlier time period, so it would not be of particular interest to test it here.

abroad.

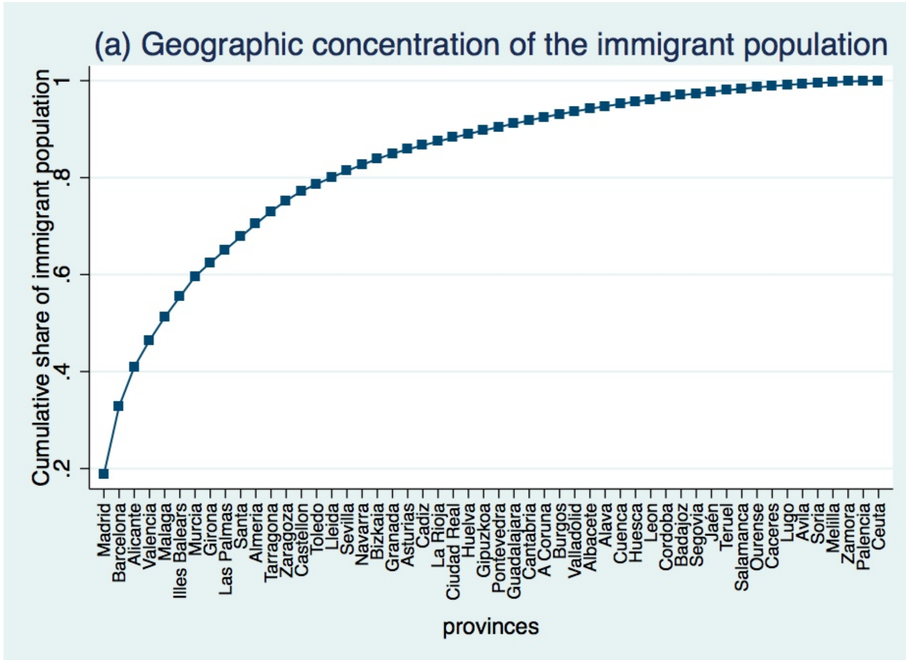
Murat and Pistoiesi [2009] and Flisi and Murat [2011] have argued that the arguments on the immigration-trade link can be reversed and applied to the cases of emigrant networks in countries such as Italy and Spain that have long emigration experiences. These authors, using country-level data on Italy and Spain, find supporting evidence to the hypothesis that emigrants facilitate exports to their destination countries through the same information, enforcement and taste effects described above; they also find the emigrants' effect on trade to be stronger than that of immigrants. However, their failure to include fixed effects at the country level or any other controls for multilateral resistance makes them run into the "gold medal mistake" of gravity estimation described by Baldwin and Taglioni [2007].

2.3 The unit of observation: province units

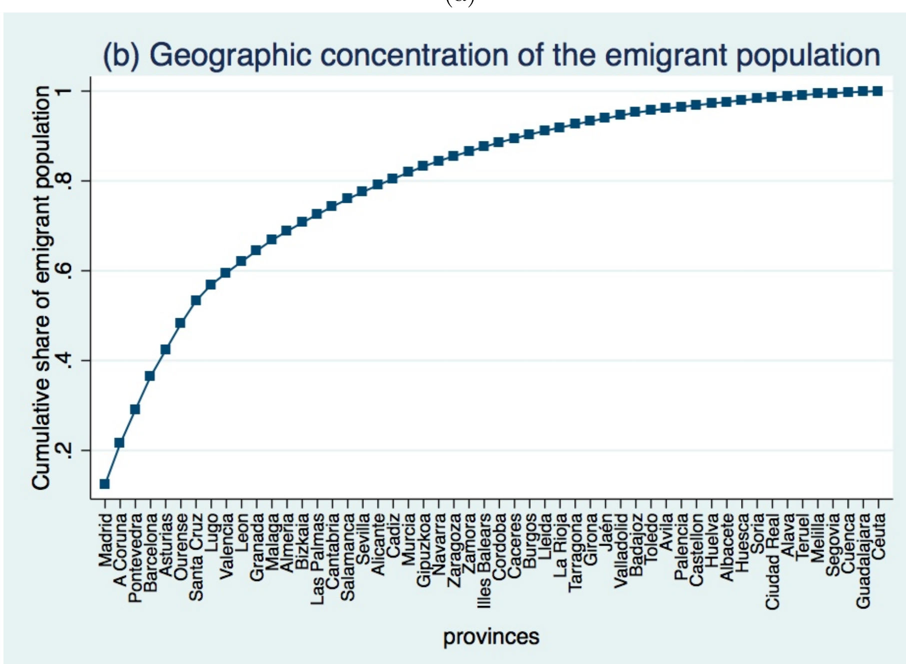
Looking at the effect of immigration on trade, one key issue to determine is the unit of observation. In most countries, the immigrant population is quite unevenly distributed. In Spain, this heterogeneity is substantial: Fig. 2 gives a hint on the geographic concentration of the immigrant and emigrant population in Spanish provinces. As panel 2a shows, 60% of the immigrant population is concentrated in seven of the 52 Spanish provinces: Madrid, Barcelona, Alicante, Valencia, Malaga, Islas Baleares and Murcia. Eight "intermediate" provinces account for a 20% of the immigrant population and the remaining 37 do not add up to more than 20%. The emigrant population (panel 2b) is also mostly originating from a few provinces: Madrid,A Coruña, Pontevedra, Barcelona, Asturias, Ourense, Santa Cruz and Lugo account for almost 57% of the Spanish expatriates.

Taking national units of observation would lead to imply an unrealistically homogeneous geographic distribution of the migrant population over the nation. Furthermore, the migration effect on trade depends on networks of individuals, families and enterprises in which proximity plays a crucial role [Rauch, 1999]. For this reason, in this paper I will follow a relevant strand of the migration-trade link literature that takes sub-national units of analysis such as provinces, federal states and departments Wagner et al., 2002, Herander and Saavedra, 2005, Dunlevy, 2006, Bandyopadhyay et al., 2008, Peri and Requena-Silvente, 2010, Briant et al., 2009, Bratti et al., 2012). While the analysis of immigration based on these units may be complicated by the high mobility of immigrants across sub-national borders (Borjas [2003] and the availability of data on migrants' characteristics radically shrinks at this disaggregation level, properly accounting for the geographic distribution of the immigrant population is deemed more important. Furthermore, for the available data, the variation and number of observations increase sizably.

Applying gravity theory to subnational units is rather straightforward, if one applies the argument in Head and Mayer [2014]: the exporting capacity S_i in equation 6 and the



(a)



(b)

Figure 2: Geographic concentration of the immigrant and emigrant population by province. Source: Own elaboration on INE data.

expenditure-weighted market access Ω_i in equation 4 need not be equal across provinces; they may vary because of, for instance, local-specific comparative advantages. Thus, by cross-sectional data, the multilateral resistance term on the sub-national unit is to be substituted by, as relevant, province, department or county effects; in the case of Spanish data, where exports are recorded by province of origin and country of destination, the argument in Baldwin and Taglioni [2007] about the use of time-varying dummies and pair effects to appropriately take account of the multilateral resistance term and of unobservables in bilateral trade determinants leads the micro-founded set of effects to include province-time, country-time and province-country effects. However, with the important exception of Briant et al. [2009] who explicitly take account of the heterogeneity at the département level, empirical studies adopting sub-national units are often omitting an explicit reflection on whether the provincial units are to be treated as fully-fledged trading partners. Instead, they seem to assume that the size disproportion between the average country and the average province justifies neglecting the economic "mass" of the subnational unit and its multilateral resistance. Wagner et al. [2002], for instance, do not add province dummies in their pooled cross-section on Canadian provinces exports arguing that "the special relationships that affect both trade and immigration likely occur politically at the national level" (p. 514-515); Bandyopadhyay et al. [2008] and Peri and Requena-Silvente [2010], using panel data on the exports of respectively US states and Spanish provinces include province-country dummies and country-time dummies but arbitrarily impose a restriction on exporter-time effects to be all the same, without a specific reference to theory. Peri and Requena-Silvente justify this by assuming that the wage levels across provinces will be the same and will thus be included in simple time dummies. Considering that within-country wage differences may be substantial and that the additional computational burden from including exporter-time effects is limited, it seems reasonable to include this set of dummies. As I will show below, the evidence on Spanish provinces shows that the province-time effects are jointly significantly different from zero; yet they do not make a big difference in terms of the point estimates of the variables of interest. Bratti et al. [2012], on Italian province-level data, propose a mixed approach where exporter-time and pair effects are included at the regional instead of the provincial level (as region-time and region-country effects), but control for province-specific income and for bilateral distance between provinces and countries of destination.

The trade cost term on sub-national units includes all the bilateral-specific factors that are affecting trade between a given province and a given country, including distance, borders, language commonality, trade agreements, (which clearly do not change from one province to another), migration of people from one country to a province (the immigration variable of this study), and from one province to one country (the emigration variable). Immigrants and emigrants would affect trade through the same information, enforcement, trust and demand mechanisms described above; clearly, however, different locations could benefit differently from the trade-facilitation role of migrants based on the bilateral specificities, accounted for by the pair effect. In one of the specifications, I will address whether the elasticity of trade to migration differs by region[6].

In addition, the use of sub-national units provides some insights on whether the diffusion of information within immigrants' networks has a geographic dimension: Herander and Saavedra [2005] and Bratti et al. [2012] find that province exports, on the whole, are much more strongly affected by the stocks of immigrants residing within the province than by those residing outside the province, supporting their hypothesis that geographic proximity within the network of immigrants matters to trade. Herander and Saavedra [2005] also find that provinces with a high concentration of immigrants from a certain nationality affect the exports of provinces with lower levels of concentration, suggesting potential information flows from high-concentration to low-concentration provinces.

2.4 Review of the empirical works on the migration-trade link on sub-national units

In this section I will briefly review the empirical works on the migration-trade link that use subnational units and are of highest relevance for this paper[7].

Wagner et al. [2002] run an extensive review of the empirical work on the migration-trade link and analyse the trade of 5 Canadian provinces with 160 countries in a pooled cross-section running from 1992 to 1995. Based on the consideration that the inclusion of fixed effects increases the "noise-to-information ratio" and biases estimates downward [Griliches, 1986], they opt for the inclusion of fixed effects at the country level but exploit the cross-sectional variation at the province level. This allows them to control for time-invariant factors that simultaneously affect trade and immigration and that could bias the estimates of the immigration elasticities. This applies in particular to language commonality (including minority languages) which they account for through a variable that captures the probability that two randomly chosen individuals within a trading pair will speak the same language. In their review, they note that, with the important exceptions of Gould [1994] and Rauch and Trinidade [2002], most studies on the migration-trade link assume constant elasticity of trade to immigration: this implies that a given increase in the immigrant stock will lead to the same increase in bilateral trade regardless of the size of the immigration stock. Noting a lack of theoretical foundations for more realistic diminishing returns on trade from immigration, they develop a "random encounter" model. This is based on the

[6]in Table A.17 in the Appendix, I also report the results of a model where the elasticity of trade is allowed to vary by province; however, the province-specific estimates rely on a relatively limited number of observations and are to be interpreted carefully.

[7]Other works on the migration-trade link that use subnational units are Co et al. [2004], Bardhan and Guhathakurta [2004] on US state exports and Combes et al. [2005]on the intra-national trade of 93 French Departments.

idea that there are two types of trading opportunities: the "hard" and the "easy" trading opportunities; the facilitation of a migrant is required to grasp the "hard" opportunities, while the easy ones get exhausted. Because the "hard" opportunities require skills and knowledge, a given immigrant has a probability p of being able to facilitate the exchange. Thus, the probability that at least one immigrant is able to facilitate the exchange is

$$1 - (1 - p)^{\mathrm{immi}_{ni}} \tag{7}$$

, which implies diminishing returns to immigration. The authors estimate their models through the Heckman procedure and by MLE, and find evidence in support of the information, enforcement and taste effects of immigrants on trade, and of diminishing returns to immigration; their results show that the effect of language commonality is not significant when immigration controls are included, leading them to conclude that the immigration effect goes beyond a language effect.

Herander and Saavedra [2005] focus on US states trade with a focus on identifying the role of geographic proximity in immigrants' networks in promoting trade. Their data is a pooled cross section on the exports of 50 US states and 36 countries, over 1993-1996, where they admit they do not exploit the time dimension because of their interest in time-invariant factors. They include controls for ancestry of the immigrant population, skills, length of stay and quality of institutions. The latter is obtained by interacting the immigration variable with a rule of law index based on Kaufmann et al. [1999]. Because of the high number of zeros in their dependent variable, they use a threshold-Tobit model based on Eaton and Tamura [1994], where the dependent variable is $ln(X_{ni} + \delta)$ where δ is a threshold to be estimated; they also compare their estimates with Powell [1984]'s Least Absolute Deviation estimator to ensure spatial and serial correlation do not affect the results. As mentioned above, they find that province exports, on the whole, are much more strongly affected by the stocks of immigrants residing within the province than by those residing outside the province, supporting their hypothesis that geographic proximity within the network of immigrants matters to trade. Another finding is that provinces with a high concentration of immigrants from a certain nationality seem to have an effect on the exports of provinces with lower levels of concentration, suggesting informational spillover effects. Their failure to include fixed effects makes them run into the "gold medal mistake", while the effect of time-invariant variables could, according to Cheng and Wall [2005], have been accounted for by regressing the estimated dyadic effects on the time-invariant variables of interest.

Dunlevy [2006] focuses on cross-sectional data on the trade of 50 US states trade to 87 countries in 1990 and, motivated by the high number of zeros in the dependent variable, applies a Tobit model where the dependent variable is the log of exports plus one. His focus is on whether immigration affects more strongly the trade of US states with weaker institutions and more corruption, hypothesizing that immigrants' "enforcement effect" substitutes for the weak institutional setting. He finds support in favour of this hypotheses and of a few other corollaries and includes in some specifications both country and province effects.

Briant et al. [2009] analyse a pooled cross-section on the trade of 94 French *départements* with 100 countries in 1999-2001. They explicitly account for the sub-national heterogeneity by adding fixed effects at the *département* level as well as country fixed effects. They also account for the potential heteroskedasticity bias noted by Santos-Silva and Tenreyro [2006] in log-linearised equations and run their estimates on the immigration effect on trade using the two-step Negative Binomial procedure suggested in Head et al. [2010], besides OLS. They address the potential endogeneity of trade to immigration by instrumenting the immigration variable with the lagged values of the variables in 1990, 1982 and 1975 finding, however, that estimates are not importantly affected. Running separate regressions on the trade of simple and complex products, based on Rauch [1999] classification, and interacting their immigration variable with a rule of law index, on the export side they find that immigrants matter to exporting all kinds of products, and more so when institutions are weak; on the import side they find that immigrants matter to the trade of simple goods only when institutions are weak, and to the trade of complex goods by higher institutional quality.

Bandyopadhyay et al. [2008] are the first to employ panel data with subnational units, using a panel on 51 US states exports to 29 countries in 1990 and 2000. They draw on Cheng and Wall [2005] argument in favour of bilateral pair-specific fixed effects to appropriately account for unobserved heterogeneity in gravity models, and apply their methodology to sub-national units. They run panel regressions with pair-specific fixed effects and country-time dummies. Their main contribution is in terms of showing that country-specific elasticities of trade to immigration may substantially depart from the elasticity measured for the full sample and are significant only for a subset of countries; this leads the authors to infer that the impact of immigration on trade differs by ethnicity. As mentioned, however, they omit a full reflection on the multilateral resistance. Rather than as the way to include multilateral resistance controls in panel data recommended by Baldwin and Taglioni [2007], fixed effects are included as a more handy way to account for heterogeneity and country-specific macroeconomic changes; the inclusion of state-time effects is completely neglected.

Peri and Requena-Silvente [2010] use a panel on province-level exports of 50 Spanish provinces to 77 countries in the 1995-2007 period. They derive from the Chaney [2008] model a specification that is similar to the one used by Bandyopadhyay et al. [2008]; drawing on Chaney [2008], however, they are able to test hypotheses on the effect of immigrants on the intensive and extensive margins of trade. They find that the immigration effect is concentrated on the extensive margin of trade: according to their results, immigrants contribute to lowering the fixed costs of opening new business ventures in their home countries to a much greater extent than they increase the average value per transaction. Peri and

Requena-Silvente [2010] also apply instrumental variables regression to exclude endogeneity from trade to immigration and, as Briant et al. [2009], find a minor bias derived from endogeneity. Within their robustness checks, they find that the immigration effect is concentrated in the provinces with a high share of immigrants over the total population and is stronger in more recent years. While they briefly refer to the application of Poisson estimation to address the potential heteroskedasticity bias in log-linearised models, they do not specify the underlying model. As mentioned above, they assume equal wage levels across provinces and do not include province-time dummies.

Bratti et al. [2012] present a peculiar model which integrates the use of province-level data on trade and immigration from Italy with the inclusion of region-time and country-region effects instead of province-time and country-province effects. This approach has the advantages of allowing the inclusion of province-level determinants such as income and distance while reducing the saturation of the model and thus leaving more scope for the variation in immigration and trade to be explained by the model. However, the arguments for adopting this approach are not fully grounded from the theoretical point of view - again, while income levels and thus expenditures and production are allowed to vary in their model, the average market access and level of competition, i.e. the multilateral resistance terms are implicitly assumed to be the same within the same region without an explicit comparison with the results of the model that uses the theory consistent set of dummies. Contributions of the paper also include the use of an IV approach similarly to Briant et al. [2009] and Peri and Requena-Silvente [2010] and the analysis, which draws on Herander and Saavedra [2005], of the effects of distance on the networks effectiveness in promoting trade.

2.5 Hypotheses

This paper aims to integrate the insights coming from the country-level study by Murat and Pistoiesi [2009] and Flisi and Murat [2011] in terms of the trade-facilitating effect of emigrant networks with those of the literature on sub-national units. It takes the case of Spain because of its peculiar relevance to the problem: it is a country with a long-lasting emigration hystory coupled with recent and booming immigration[8]. It aims at testing a number of hypotheses on the effect of immigration and emigration stocks on the exports of sub-national units with a theoretically and econometrically consistent methodology. This implies appropriately taking into account the multilateral resistance term as well as addressing methodological problems which stand at the "frontiers of gravity research" [Head and Mayer, 2014] and are outlined in Section 4.

The first basic hypothesis is that *both emigrants and immigrants have a significant effect*

[8]Another reason to lead this analysis on the Spanish case is the public availability of data on the main variables of interest, especially the trade of provinces and the immigrant and emigrant stocks.

on the trade of Spanish provinces (Hp 1.1). This hypothesis will be tested along with two "corollaries" derived from the literature, that *immigrants' and emigrants' effect on trade is stronger with more culturally distant (Hp 1.2) and more institutionally distant countries (Hp 1.3).*

The second set of hypotheses relates to subnational heterogeneity. First, because there is no reason to think that at least some of the trade costs and economic masses vary at the province level as well as the country level, I hypothesize that *province-level heterogeneity in the multilateral resistance term has an impact on trade (Hp. 2.1).* Second, I will use a different estimation method to test Herander and Saavedra [2005]'s hypothesis that *local networks of immigrants (and emigrants) from the same province have a stronger effect on the trade of provinces than more distant networks (Hp 2.2).* Third, I will test, using a method which is very similar to that used by Bandyopadhyay et al. [2008] to test ethnic heterogeneity in the migration effect, whether *there is heterogeneity among regions in the immigrants' and emigrants' effects (Hp 2.3).*

The third set of hypotheses looks at non-constancy in the elasticity of trade to immigration and emigration. First, I hypothesize that *the effect of immigration and emigration is differentiated by the shares of immigrants in each province (Hp 3.1).* This has been found to be the case for immigration by Peri and Requena-Silvente [2010] on the 1995-2007 period. Second, I test the hypothesis that *the immigration and emigration effect is non-constant (Hp 3.2).* This is to test the non-linearity hypothesis proposed by Gould [1994] and Wagner et al. [2002]: before they can exert an effect on trade, a certain "mass" has to be reached by immigrants; after a given level, the returns from immigration to trade are diminishing.

3 Data

The database used for the empirical analysis is a balanced panel based on export data about 50 Spanish provinces[9] (NUTS 3) and 65 countries over 5 years (2006-2010). The selection of the countries is driven by the availability of province-level data on immigrant and emigrant stocks, and by whether their share on total Spanish exports is at least 0,1% every year. Overall, the selected countries account for more than 91% of total Spanish international trade for each year of the panel (see table A.15 in the Appendix for the complete list of countries).

The interpretation of the migration variables is subject to some caveats. Following the literature on the migration-trade link, throughout this paper I refer to "immigration" in a province as the stock of residents in that province who hold a non-Spanish citizenship

[9]The provinces of Ceuta and Melilla are excluded for data availability reasons.

(data on immigration in Spain are from the National Statistics Institute, INE; see Table A.16 in the Appendix for a list of the main data sources). As it is common in migration studies, this is an imperfect measure of immigration[10] since it neglects the portion of foreign-born people that have acquired the nationality of the host country. It also only refers to formally residing people, neglecting undocumented immigrants, and it furthermore neglects the intra-national mobility that is not registered in changes of residence. Similarly, the emigration variable used here is imperfect as it refers to the stock of people that have moved their residence outside Spain but are still recorded in the election registries in Spain (as in Flisi and Murat [2011], data originate from the CERA, *Censo Electoral de Residentes Ausentes*). These data are not informative as to the country of birth of these emigrants, and thus in principle one cannot distinguish return migrants from the native Spanish diaspora: only the portion of former residents in Spain who have expatriated and who still vote in Spain is represented. Yet, maintaining one's voting rights in Spain implies the persistence of strong ties to Spain. It is thus relatively safe to assume these data to more closely reflect the dynamics of the Spanish emigrant population than the dynamics of return migration. Due to often circular migration ties between Spain and several Latin-American countries, however, the results on the emigration variable should be interpreted with particular care [11]. Furthermore, neither immigration nor emigration data at the province level allow distinguishing between economically active and inactive migrants; consequently, it cannot be elicited from these data whether migrants' contribution to trade is "active", e.g. as entrepreneurs, intermediaries or labour force who sell primarily to their home country, or whether it is an "indirect" one that goes through familiarization and trust-building in the destination context with the home countries. This is a consequence of the sub-national level of analysis: at the NUTS 3 level, the availability of detailed data on immigrants' characteristics, especially on skills, employment status and length of stay is severely constrained: the results of the Labour Force Survey are only considered as reliable at the NUTS 2 level.

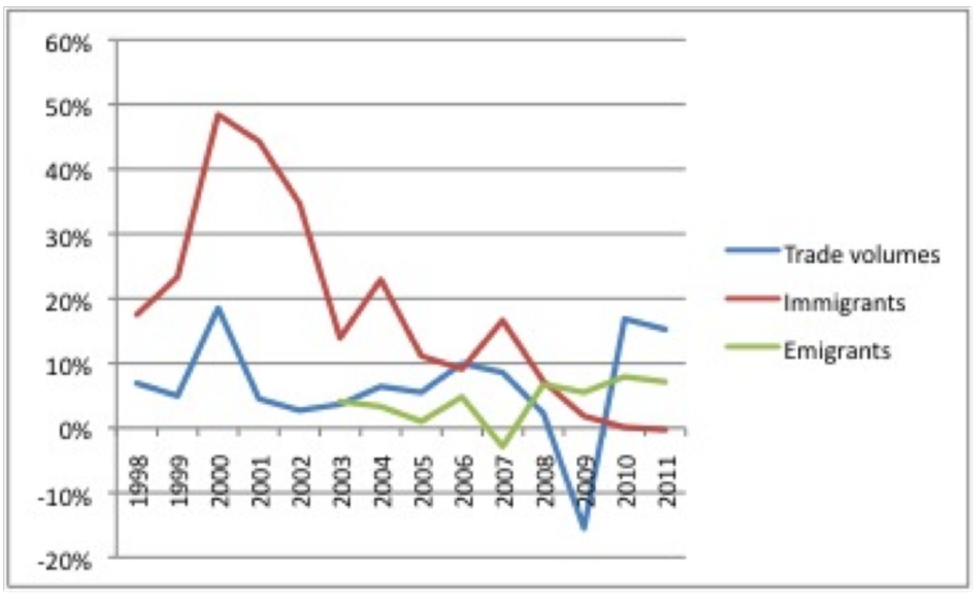
Fig. 3 shows the growth rates of trade volumes, immigrant stocks and emigrant stocks over the most recent period. For data availability reasons of the emigrant variable, this period includes the period of the burst of the global financial crisis. On the whole, the three variables have been growing over the period, with exports growing at an average

[10] "Immigration denotes the action by which a person establishes his or her usual residence in the territory of a Member State for a period that is, or is expected to be, of at least 12 months, having previously been usually resident in another Member State or a third country"; "Emigration denotes the action by which a person, having previously been usually resident in the territory of a Member State, ceases to have his or her usual residence in that Member State for a period that is, or is expected to be, of at least 12 months" (Eurostat definitions: `http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/migr_flow_esms.htm/unit_measure`).

[11] The available INE data on residential variations to foreign countries report that from 2004 to 2011 the percentage of Spanish residents of Spanish nationality changing their residence to foreign countries has been less than 25% of total residential variations in all years but 2005 (when it was 28,4%).

rate of 6.46%, emigration stocks growing at an average rate of 4.17%, and immigration stocks booming at an average rate of 17,9%. However, the 2008-2009 crisis period has severely affected both exports and immigrant stocks. While the exports have rapidly recovered, this period has brought the yearly growth of immigration stocks to stagnation. On the contrary, emigrant stocks, which seem to roughly follow an opposite pattern than immigrant stocks, have been growing faster since the crisis years on. The extremely high levels of unemployment associated with the crisis in Spain are probably responsible for these changes.

Figure 3: Growth rates in trade volumes, immigrant and emigrant stocks in Spain (country-level data, 1997-2011). Source: own elaboration on Datacomex and INE data.



The correlation between the immigration and emigration variables is 0.10, so the two variables can be assumed to portray quite different phenomena. Indeed the distribution of immigrants and emigrants across provinces follows quite distinct, in some cases opposing, patterns: where the immigrant share over total population in a given province is high, the corresponding share of emigrants is often low (see Table A.17 in the Appendix, reporting data about immigrants and emigrants distribution across provinces in 2010). As well as the concentration of foreign population seen in section 2.3, also the share of residents with a foreign nationality shows substantial variation across provinces. In 18 out of 52 provinces the share of foreigners was above the national average in 2010, reaching above 20% in four provinces (Alicante, Islas Baleares, Almería, Girona); in 15 provinces the share was below 5%; it lagged below 4% in Cádiz, Cáceres, Badajoz, A Coruña, Córdoba and Jaén.

As to the origin and destination countries, Fig. 4 shows the subset of the most frequent origin countries for immigrants and destination countries for emigrants. While in both cases many are OECD and Spanish-speaking countries, the ranking is quite different. Romania and Morocco are the most frequent origin countries for immigrants (respectively accounting for 14.5% and 13.1% of the immigrant population), while Argentina is by far the most popular destination country for emigration, hosting 22% of the Spanish expatriates (See also Table A.18 in the appendix for some information about the profile of the migrants).

Figure 4: Top 15 origin countries of immigrants and destination countries of emigrants in Spain, 2010. Source: Own elaboration on INE data.

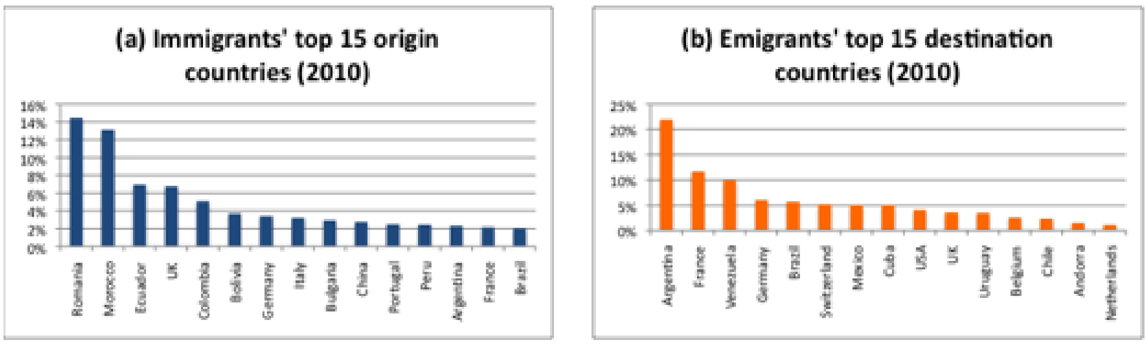


Table 1 reports the summary statistics for main variables of interest of this paper: the value of the exports from province i to country n in year t (X_{nit}), the stock of immigrants from country n living in province i at year t ($Immi_{nit}$) and the stock of emigrants from province i living in country n at year t (Emi_{nit}). Included are also province i gross product in year t (Y_{it}), country n GDP in year t (Y_{nt}), province i population in year t (N_{it}), country n population in year t (N_{nt}). The extent of the correlation is higher between exports and each of the migration variables (the correlation of exports is respectively 0.15 with immigration stocks and 0.24 with emigration stocks[12]), than between immigration and emigration stocks. The correlation of the migration variables with the province income is relatively high - the correlation between province income and immigration is 0.33; between province income and emigration it is 0.17. It is, instead very low with the country income variable: between emigration and country income the correlation is 0.08; it goes to almost zero (0.001) between immigration and country income. This suggests that probably the correlation is to be found at the local rather than at the country-level and is another element in favour of the use of subnational units.

As the summary statistics in Table 1 show, the variation in the data is high. The export,

[12]The correlations between the log of the exports and the log of the two variables are respectively 0.39 and 0.45.

Table 1: Summary statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
X_{nit}	value of the exports from province i to country n in year t	15195	53667.32	243745	.00	7208594
$Immi_{nit}$	stock of immigrants from country n living in province i at year t	15195	1474.95	7021.12	1	210823
Emi_{nit}	stock of emigrants from province i living in country n at year t	15195	389.32	1880.05	1	46545
Y_{it}	province i gross product in year t	15195	2.18 e+07	3.23 e+07	1927919	1.93 e+08
X_{nt}	country n GDP in year t	15195	8.41 e+08	1.99 e+09	2823504	1.45 e+10
N_{it}	province i population in year t	15195	939.34	1134.77	91.90	6358.60
N_{jt}	country n population in year t	15195	80714.15	221064.50	79.87	1341414

income and migration data are typically highly concentrated around the small values of the distributions, with long right tails. This is typical in trade data, combining data about pairs that may be radically different in terms of economic sizes (and, consequently, in terms of their push/pull factors for immigration and emigration). This variance is usually reflected in heteroskedasticity, which in turn can seriously bias the estimates when they are based on logs, as noted by Santos-Silva and Tenreyro [2006] (see section 4). The migration variables, especially the emigration variable, are quite zero inflated: the share of zeros out of the total number of observations is 4,09% for $Immi_{nit}$ and 13,45% for Emi_{nit}. Instead, because of the data selection procedure described above, the data on exports only have a negligible number of zeros: the unit of exports is thousands of euros, and the share of province exports to partner countries below 1 mile of euros is limited to 0,24% of total observations.

Figure 5 brings together the data on trade and migration for the province of Madrid in 2010.

Running a similar exercise as in Fig. 1, in Fig. 5 I plotted, for Madrid's OECD trade partners, the export/GDP ratio to each with the distance-weighted immigrant and emigrant stocks from each country. The rationale for taking distance-weighted immigration and emigration stocks is that migration typically decreases with distance. The rationale for taking OECD countries only is to achieve a certain degree of institutional similarity which can be expected to reduce the omitted variables bias in the relation. The relationship appears positive and the immigrants simple regression line is steeper than the emigrants'.

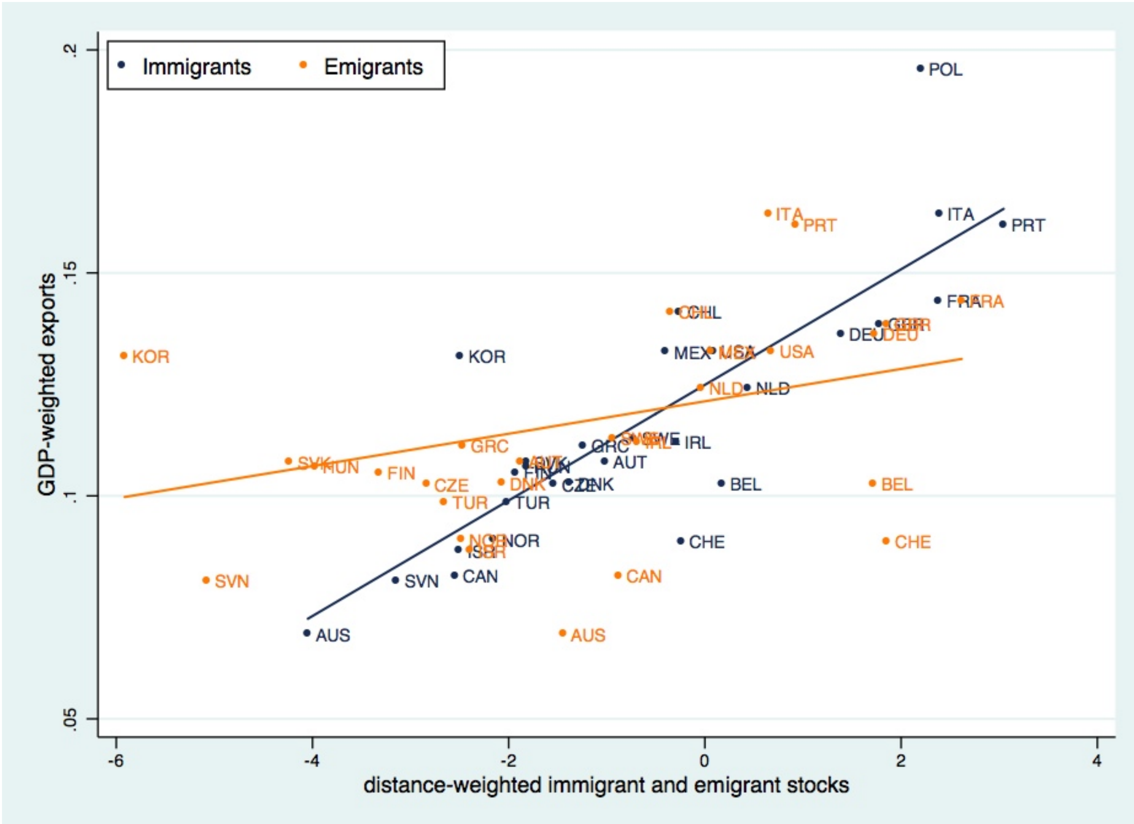


Figure 5: The migration-trade link: province of Madrid and OECD countries, 2010. The graph shows the relationship between GDP-weighted exports and the distance-weighted immigrants and emigrant stocks. Own elaboration on Datacomex, INE, CERA and FMI data.

The picture is similar when restricting the analysis to EU countries only. This purely descriptive result motivates a more rigorous econometric analysis of the relationship.

4 Econometric Strategy

4.1 Main econometric issues in gravity models

Briant et al. [2009] identify three econometric issues to be addressed for consistently estimating gravity models: zeros, heteroskedasticty and endogeneity. This partially contributes to explain the variety of estimation methods outlined in the review above.

4.1.1 Zero trade flows

The issue of zero trade flows is a concern for the international trade literature in general [Helpman et al., 2008] and is particularly relevant for the trade of sub-national units, where small flows are common and zero flows are frequently observed. The problem of zero flows arises from the log linearization in gravity models, due to the indeterminacy of the log of zero. Simple OLS estimates are thus only possible on non-zero observations. The censoring at zero introduces a bias in the simple OLS estimator and the model is more consistently estimated in levels through Tobit models or models of the Poisson family, as argued by Santos-Silva and Tenreyro [2006] and Eaton and Kortum [2002]. Eaton and Tamura [1994] threshold Tobit model is one of the first applications of the Tobit estimator to gravity models which has been applied to the migration-trade link literature by Herander and Saavedra [2005] and by Dunlevy [2006]; Wagner et al. [2002] use the Heckman procedure. As to my data, the sample selection dictated by the need to ensure that data are available on the migration variables causes the issue of zero trade to be negligible, limited to only 0.24% of the observations. Hence, the issue of zero trade will not be addressed here.

4.1.2 Heteroskedasticity and functional form of the errors

Santos-Silva and Tenreyro [2006], drawing on Manning and Mullahy [2001], note that the usual procedure of log-linearizing the gravity equation and estimating it by OLS could introduce a bias in the estimates when the error term is heteroskedastic. Because the expected value of the logarithm of a random variable depends both on its mean and on the higher-order moments of its distribution, in presence of heteroskedasticity it cannot be assured that the unconditional mean of the log of the errors is zero, and that their conditional mean is independent from the covariates. In practice, the variation in trade data typically increases by higher trade flows; higher trade flows are in general associated with greater economic size and lower bilateral trade costs, i.e. the gravity model covariates. While this poses no problem in standard OLS estimation, by log-linearisation the variance enters the determination of the expected value of the error term. Thus, a violation of the homoskedasticity assumption will in general lead to the fact that expected value of the log-linearized error depends on the covariates, leading to inconsistent OLS estimates. The bias in the OLS estimation is partially reduced by fixed effects. Santos-Silva and Tenreyro [2006] compare OLS, Gamma regression, non-linear least squares (NLS) and Poisson pseudo-maximum likelihood estimators in a cross-sectional context. They argue in favour of the use of a Poisson pseudo-maximum likelihood estimator as it is more consistent than OLS and more efficient than NLS. Because the Poisson estimator allows accounting for the zeros in the dependent variable and is consistent even by variance misspecification [Wooldridge, 2002], the authors prefer the Poisson Pseudo-Maximum-Likelihood Estimator

(PPML) to other count models such as the Negative Binomial[13].

The Santos-Silva and Tenreyro argument to make the PPML the new "workhorse" for gravity models in the presence of heteroskedasticity has been recently challenged by Head and Mayer [2014] in their handbook chapter. The selection of the appropriate estimation method based on considerations on the functional form of the errors lies in what the authors call the "frontiers of gravity research". The authors compare the moment conditions of the PPML and of the OLS with those of the Gamma PML and note that the OLS and Gamma PML pursue a similar objective: in the case of the Gamma, to set to zero the sum of the percent deviations of actual trade from predicted trade; in the case of OLS, to set to zero the sum of the log deviations of actual from predicted trade. The two are approximately equal (p.40). The Poisson PML moment conditions, instead, set to zero the *level* deviations of the actual from predicted trade. For this reason, Poisson PML gives more weight to larger observations compared to the other two estimators. The Gamma regression, like the Poisson PMLE, delivers consistent estimates regardless of the functional form of the errors as long as the conditional mean is correctly specified, because it also measures the dependent variable in levels. The relative efficiency of the two estimators depends on the assumptions made about the variance of the errors. Drawing on Santos-Silva and Tenreyro [2006] and on Manning and Mullahy [2001], Head and Mayer [2014] propose a general relation linking the variance to the conditional mean which contains the different cases:

$$\mathrm{var}[X_{ni}|\mathbf{z}_{ni}] = h\mathrm{E}[X_{ni}|\mathbf{z}_{ni}]^{\lambda} \tag{8}$$

The Poisson PML is the most efficient estimator by $\lambda = 1$, corrisponding to a constant variance-to-mean ratio, a generalization of the Poisson assumption of $h = \lambda = 1$. The homoskedastic OLS estimator in logs is the MLE by $\lambda = 2$, which corresponds to a constant coefficient of variation. By heteroskedasticity and $\lambda = 2$, the consistent and efficient estimator is the Gamma PML[14].Both the Poisson and the Gamma PML remain consistent

[13]Santos-Silva and Tenreyro also argue against the use of the Negative Binomial in their "Log of gravity" webpage,`http://privatewww.essex.ac.uk/~jmcss/LGW.html`:

> "The estimator proposed in the Log of Gravity is simply a weighted non-linear least squares estimator. It turns out that with the proposed weights, the first-order conditions for this estimator are identical to those of the Poisson pseudo-maximum likelihood regression. Therefore, the fact that we recommend the used of a count data estimator for the gravity equation is just a fortunate coincidence that allows the use of a well-known regression method which is widely available is econometric and statistics software. (...) both the negative-binomial and the zero-inflated regression models have the important drawback of not being invariant to the scale of the dependent variable. That is, measuring trade in dollars or in thousands of dollars will lead to different estimates of the elasticities of interest!"

Head and Mayer [2014], too, urge researchers to "resist the siren song of Negative Binomial", mainly because of the problem that the estimates depend on the unit of measurement in the dependent variable (Boulhol and Bosquet [2012]).

[14]Another possible option is the heteroskedastic-retransformed OLS estimator (Manning and Mullahy [2001], which is however not practicable in my case.

by $h > 1$, i.e. by overdispersion [Wooldridge, 2002]. Equation (8) can be log-transformed and applied to the sample counterparts of the mean and variance parameters to become the Manning and Mullahy [2001] test on the underlying functional form of the errors, or Santos-Silva and Tenreyro [2006]'s "Park-type test"; Head and Mayer [2014] dub it "MaMu test":

$$\ln\hat{\epsilon}^2_{ni} = \text{constant} + \lambda\widehat{\ln X}_{ni} \tag{9}$$

Based on the results of a series of Monte Carlo simulations, Head and Mayer [2014] confirm the robustness of the Poisson and Gamma PML to different functional form of the errors. However, they also find some evidence against the Poisson as the new "workhorse" for gravity equations: by mis-specification of the conditional mean, the Poisson PML results severely biased due to the higher weight given to larger observations, while the OLS on logs and the Gamma PML are consistent by sufficiently large sample sizes. The kind of model mis-specification that they introduce is of particular interest for empirical works on the migration-trade link: they assume non-constancy in the variable of interest. As seen above, while the large majority of the recent works assume constant elasticity of trade to migration, earlier works explored the hypothesis of diminishing returns to migration [Gould, 1994, Rauch and Trinidade, 2002, Wagner et al., 2002].

Hence, none of these estimation methods can be considered as a single "workhorse". Instead, the consistent and efficient estimator in relation to the specific issues posed by a given dataset is more likely to be identified from the comparison of the three estimation methods. Head and Mayer [2014] identify a few simple selection criteria (p.44):

1. If the three give similar results, the model appears well specified, with approximately log-normal errors and constant variance. In this case, the OLS is the MLE.

2. If the Poisson and Gamma PML estimates are similar to each other and different from the OLS, heteroskedasticity should be considered as a problem and the OLS is to be considered as inconsistent.

3. If the Gamma PML and OLS coefficients are similar and the Poisson estimates are smaller in absolute values, there are two possible interpretations: (a) By small samples, this may be due to small sample bias in either of the PML estimators. If the $\hat{\lambda}$ derived from the MaMu test in Equation (**??**) is significantly below 2, Poisson should be preferred. (b) By large samples, significantly divergent estimates may a signal non-constant elasticities and model mis-specification.

In what follows, I will apply this approach. I will run the three estimators (OLS, Poisson and Gamma) on my data, and compare their results, as outlined more in details in section 4.2.

4.1.3 Reverse Causality

According to Briant et al. [2009], the third econometric issue affecting trade studies is endogeneity, which may arise from reverse causality or omitted variables.

As to reverse causality, Briant et al. [2009] Peri and Requena-Silvente [2010] and Bratti et al. [2012], using sub-national trade and migration data on France, Spain and Italy, apply instrumental variables approaches; their IV estimates, mainly based on imputed values of immigration drawn by the distribution of immigrants in the nineties, confirm the estimates obtained by OLS. On the other hand, as argumented by Gould [1994] and Briant et al. [2009]it is unlikely that the preferential bilateral links between two countries are simultaneously causing trade and migration flows: migration flows, are mainly driven by family reunification motives, wage and income differentials and similarity in languages (see also [de Haas, 2010]). Therefore, I will assume that reverse causality is not a problem in my data and I will not address it specifically in my estimates.

4.1.4 Omitted variables and heterogeneity

Another possible ground for the violation of the assumption of independence of the errors may derive from omitted variables.

In the case of the gravity models, a large component of the multilateral resistance terms, as well as of the bilateral trade determinants, are known to be unobservables (Anderson and van Wincoop, 2003, Baldwin and Taglioni, 2007; see also section 2.1). By panel data, the time-varying component of the multilateral resistance terms should be accounted for by time-varying effects; the correlation between unobservable bilateral trade determinants and included bilateral trade determinants (migration variables, in this case) should be accounted for by dyadic time-invariant effects [Baldwin and Taglioni, 2007].

Additional unobservables may be the consequence of heterogeneity at the level of the trading pair that affect trade through the bilateral trade costs channel. This is discussed, for the case of panel data, in Cheng and Wall [2005] and Bandyopadhyay et al. [2008]. These studies argue in favour of the inclusion of bilateral pair-specific fixed effects, discarding other, less computationally burdensome, simplifications deriving from arbitrary restrictions imposed to the effects (e.g. symmetric pair-specific effects, a single fixed effect for a given importer towards all exporters, a single exporter effect towards all importers).

Including the three sets of effects (importer-time, exporter-time and dyadic) implies that bilateral-specific time-invariant variables are absorbed in the trading pair effects, and that importer-specific and exporter-specific time-varying variables are absorbed in the importer-time and exporter-time effects. This leaves out for estimation only those variables that are at the same time dyadic and time-varying, such as immigration and emigration stocks.

The absorption of variables of interest in the fixed effects does not prevent a analyzing their role in determining trade flows. According to Cheng and Wall [2005], the unbiased estimation of time-invariant variables such as distance, language and institutional similarities is best performed by running a regression with the estimated pair effects as dependent variable and the time-invariant variables as regressors. This technique leads them to confirm the effect of time-invariant variables in the expected directions, in line with the findings of the gravity literature.

4.2 Econometric Model and Steps

Based on the considerations above, the optimal strategy would be to include the three sets of effects in the OLS, Poisson and Gamma estimations as suggested in section 4.1.4 and to proceed with the steps outlined in section 4.1.2 to identify the suitable model for the specific data at stake. While this poses no problem to the OLS and Poisson estimation, Gamma estimation with many dummies in short panels is problematic. Here, the estimation of the coefficients of the bilateral fixed effects relies on the time variation only; when the panel is short, the coefficient cannot be consistently estimated; and the inconsistency could in principle pass on to the estimates of the variables of interest [Cameron and Trivedi, 2009].

To the best of my knwoledge, there is no way to estimate a Gamma regression with a high number of dummies without incurring in the incidental parameters problem. On the other hand, not including the bilateral fixed effects implies incurring certainly on omitted variable bias: the bilateral fixed effects are jointly highly significant; they are supported by likelihood ratio tests, dramatically increase the R-squared and decrease the Akaike Information Criterion statistics. Thus, I will include the three sets of effects in the Gamma regression and rely on the high number of time-varying exporter and importer effects, as well as on the very high number of observations (15195) on which the migration variables can be estimated, to prevent the possible inconsistency from passing on to the variables of interest.

The basic theory-consistent model to analyse the immigrants' and emigrants' effects on the trade of Spanish provinces is the following:

$$X_{nit} = \alpha X_{nt}^{b_1} Y_{it}^{b_2} \mathrm{Immi}_{nit}^{\beta_1} \mathrm{Emi}_{nit}^{\beta_2} e^{(\boldsymbol{\gamma_1 \theta_{nt}} + \boldsymbol{\gamma_2 \omega_{it}} + \boldsymbol{\gamma_3 \eta_{ni}} + \varepsilon_{nit})} \tag{10}$$

Where:

X_{nit} = Nominal value of the exports from the Spanish province i to country n at time t;

X_{nt} = Total expenditures by country n, approximated by GDP,

Y_{it} = Total production by province i, approximated by province gross product;

Immi_{nit} = Stock of immigrants from country n living in province i at time t;

Emi_{nit} = Stock of emigrants from province i living in country n at time t;

$\boldsymbol{\theta_{nt}}$ = vector of the importer-time effects, corresponding to country-time dummies;

$\boldsymbol{\omega_{it}}$= vector of the exporter-time effects, corresponding to province-time dummies;

$\boldsymbol{\eta_{ni}}$= vector of the trading-pair specific fixed effects, corresponding to province-country dummies

The right-hand side of the equation is log-linearised in the estimation. The left-hand side of the equation is in levels for in the cases of the Poisson and Gamma PML estimation and in logs in the case of OLS estimation.

In practice, estimating X_{nt} together with importer-time effects, and Y_{it} together with exporter-time effects in the same equation gives rise to perfect multicollinearity[15].

As mentioned in section 3,to address potential non-linearities and the indeterminacy of the log of zero in the immigration and emigration variables, I add one unit to each observation and include a No-Immigrant-Dummy (NID) and a No-Emigrant-Dummy (NED) to the equation.

The log-linearised model to be estimated by OLS becomes:

$$ln(X_{nit}) = \beta_1 \ln(\text{Immi}_{nit}) + \beta_2 \ln(\text{Emi}_{nit}) + \gamma_{\mathbf{1}} \boldsymbol{\theta_{nt}} + \gamma_{\mathbf{2}} \boldsymbol{\omega_{it}} + \gamma_{\mathbf{3}} \boldsymbol{\eta_{ni}} + \varepsilon_{nit} \tag{11}$$

The model to be estimated by Poisson and Gamma PML is the following:

$$X_{nit} = \beta_1 \ln(\text{Immi}_{nit}) + \beta_2 \ln(\text{Emi}_{nit}) + \gamma_{\mathbf{1}} \boldsymbol{\theta_{nt}} + \gamma_{\mathbf{2}} \boldsymbol{\omega_{it}} + \gamma_{\mathbf{3}} \boldsymbol{\eta_{ni}} + \varepsilon_{nit} \tag{12}$$

As discussed above, I follow Head and Mayer [2014] in estimating the model with the three estimation methods, construing the whole strategy as a "robustness-exploring ensemble"[16].

[15] The analyst may actually be able to estimate log-linear versions of model 10 and to view coefficients for the income variables, as in Peri and Requena-Silvente [2010] and in Bandyopadhyay et al. [2008]. This comes however at the cost of omitting some dummies, which are dropped automatically by the software and are not fully controlled by the analyst. This implies that the estimated coefficients for the income variables actually refer to the single cases for which the dummies have been dropped and cannot be interpreted more generally.

[16] The Poisson estimates are run using the robust feature of the `xtpoisson` estimator in Stata12, which includes the features of `xtpqml`, the previous version of the command developed by Timothy Simcoe. The Gamma estimates are run using the `glm` command with the options `link(log) family(gamma)` with a series of trading pair dummies representing the three sets of effects and standard errors clustered at the pair level. In the case of OLS, too, unless otherwise specified, the standard errors are clustered at the pair level.

To draw conclusions about the consistent model, in section I will test for heteroskedasticity and learn about the functional form of the residuals through a MaMu test; I will compare the estimates based on the criteria proposed by Head and Mayer [2014] and draw conclusions on this basis. The selected model will be used to test more specific hypotheses (see section 4.3.

4.3 Operationalization of the hypotheses

The immigrants and emigrants' effect. Hypothesis 1.1 on the aggregate effect of immigrants and emigrants on trade will be tested in terms of the statistical significance of β_1 and β_2 in equations 11 and 12.

The "corollaries" about the expected stronger role of immigrants and emigrants promoting trade towards countries that are more distant institutionally (Hp 1.2) and culturally (Hp 1.3) will be tested by interacting $ln(Immi_{nit})$ and $ln(Emi_{nit})$ with, respectively, a D_{EU} and a D_{Spa} dummy. D_{EU} equals 1 for EU Member States and zero otherwise; the coefficients of its interaction with the migration variables are aimed to test the role of institutional (dis)similarity in promoting trade. D_{Spa} equals 1 when the country has Spanish as an official language and zero otherwise. Il closely mirrors remote colonial ties; the coefficient of its interaction terms with $ln(Immi_{nit})$ and $ln(Emi_{nit})$ are expected to test the role of cultural similarity in promoting the migration-trade link. A confirmation of the hypotheses that the migration effect is stronger towards more culturally and institutionally distant countries is usually read as a proof of the prevalence of the information and enforcement effects, mediated by immigrants and emigrants, on the preference effect mechanism of trade promotion, as both information and enforcement costs are likely to increase with institutional and cultural distance.

Sub-national heterogeneity. As this paper focuses on the exports side, the test for the heterogeneity on the multilateral resistance term (Hypothesis 2.1) will focus on the term Ω_i in the structural gravity model (equation 4), i.e. on the average market access of the province on all markets, which by panel data is time-varying and corresponds to ω_{it} in model 10. The papers which omit this term explicitly or implicitly assume that the average market access is the same across provinces. Hence, to verify the hypothesis of province-level heterogeneity in this term, I will run a statistical test on the basic specification to verify the joint statistical significance of the set of province-time effects; I will also estimate the mode for each estimator l with a different specification that excludes the province-time effects but includes province income (to ensure that the exclusion of the province-specific multilateral resistance term does not exclude the production term Y_i). I will compare the log-likelihood and AIC statistic[17] to conclude whether province-level heterogeneity is supported by the data, besides theory.

[17] Stata does not allow likelihood ratio tests on models with clustered and robust standard errors.

The test of hypothesis 2.2, that trade-relevant information more effectively flows through local networks of immigrants than over national-level networks, will be tested by applying the procedure in Herander and Saavedra [2005] to my data and estimation methods. Two additional variables called $ln(ImmiOut_{nit})$ and $ln(EmiOut_{nit})$ will be included in models 11 and 12. The term $ln(ImmiOut_{nit})$ represents the total stock of immigrants from country n living in provinces other than i at time t, and $ln(EmiOut_{nit})$ represents the total stock of emigrants registered in provinces other than i who had migrated to country n at time t. They are meant to represent national networks of immigrants and emigrants that extend beyond the province. A positive and significant effect of $ln(ImmiOut_{nit})$ or $ln(EmiOut_{nit})$ would suggest that the information that is relevant to promoting trade towards country n is exchanged country-wide, respectively within the group of immigrants from n and within the group of emigrants to n. A positive and significant effect of $ln(Immi_{nit})$ and $ln(Emi_{nit})$ only is instead to be interpreted as a sign that the information that is relevant for trade is predominantly exchanged within the local network of immigrants from country n living in province i; in the emigration case, a positive effect of the latter could be interpreted as a sign of localised information flows on business opportunities abroad occurring prior to the decision to leave.

Finally, Hypothesis 2.3 on the region-specific migration effects will be tested by letting β_1 and β_2 vary by province, in the same way that Bandyopadhyay et al. [2008] allowed it to vary by country. I expect to find evidence that the effect of immigrants and emigrants on trade is differentiated by regions. I will substitute the coefficients β_1 and β_2 of $ln(Immi_{nit})$ and $ln(Emi_{nit})$ in equations 11 and 12 with region-specific β_{1i} and β_{2i}.

Immigration and emigration non-constancy. Hypothesis 3.1 that the effect of immigration and emigration is differentiated by the shares of immigrants in each province will be tested by splitting the sample into provinces with a high, intermediate and low share of immigrants, and into provinces with a high, intermediate and low share of emigrants.

The share of immigrants in province i s_i^{immi} is calculated on the overall population of immigrants from any country over the total population of province i. Similarly, the share of emigrants from province i over the total population s_i^{emi} is calculated on all the Spanish expatriates from province i to any country. The values of s_i^{immi} and s_i^{emi} for the single province will be considered as high if they lie above the 33^{rd} percentile, intermediate if they are between the 34^{th} and the 67^{th} percentile and high if they are above the 67^{th} percentile of the values of the shares of immigrants in all provinces each year.

Hypothesis 3.2 of non-constant elasticity of trade to immigration cannot be directly tested by an adaptation of the methodology applied by Gould [1994] and Wagner et al. [2002] as it would imply the inclusion of fixed effects in non-linear least squares estimators[18]. This hypothesis will thus be addressed through the inclusion of non-linear terms in the model as

[18] Stata does not allow including fixed effects in nlls models

part of the exploration of the correct specification of the functional form of the conditional mean.

5 Results

5.1 Immigrants' and emigrants' effects on trade

In table 2 I compare the three estimation methods to address hypothesis 1.1: whether immigrants and emigrants have an effect on the trade of Spanish provinces. This table shows a pattern that is common to most of the estimation results in this paper. The OLS and Gamma estimates show a positive and significant effect of the immigrant stocks on trade, with magnitudes that are comparable with each other: the OLS estimates show that, by a 10% increase in the immigrant population, trade is expected to grow on average by 1.6%; according to the Gamma estimates, by the same increase in the immigrant stocks, trade will grow by 1.4%. The Poisson estimation, instead, does not portray any recognizable pattern of trade creation by either immigrants or emigrants. Instead, they show a positive and significant effect of the no-immigrant dummy, pointing at possible non-linearities in the immigration variable. The magnitude of NID is almost exactly the same in the Poisson PML and the Gamma PML regression.

Table 2: Regression results Hp 1.1 - The effect of immigrants and emigrants on the trade of Spanish provinces

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·162***	0·045	0·137***
	(0·061)	(0·048)	(0·032)
$ln(Emi_{nit})$	−0·012	0·030	0·005
	(0·044)	(0·031)	(0·026)
NID	0·176	0·217 * *	0·216***
	(0·111)	(0·108)	(0·065)
NED	0·030	0·083	0·045
	(0·067)	(0·063)	(0·040)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		
log-likelihood	−14537·1	$-1{\cdot}54e{+}07$	−143798·1
AIC	29994·2	$3{\cdot}08e{+}07$	288516·1

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

The numerosity of the observations allows excluding small sample bias. The scenario in which the two PML estimates, both supposed consistent, show different results, falls under case 3 of the options presented in section 4.1.2, i.e. one where heteroskedasticity in the OLS estimates is not too marked and the elasticitity of the variable of interest has been misspecified. The case presented in Head and Mayer [2014] is the case of a non-constant elasticity, which the Poisson PML mis-estimates because of the higher weight it gives to more sizeable observations in the trade variable.

It is important to note that the model presented here is derived from a vast literature which, with the exceptions of the earlier works by Gould [1994] and Wagner et al. [2002], routinely assumes a linear relationship between the log of the immigration stocks and the log of trade. Thus, confirming model mis-specification would bear important implications to this literature. This possibility will be analysed in section 5.2.

Reliance on the Gamma PML estimates, however, is in turn subject to some caveats. The incidental parameter problem affects the estimation of the fixed effects in the Gamma regression, making their estimates inconsistent. While the coefficients of the fixed effects are *per se* of little interest to the analyst, it is possible that this inconsistency affects the estimation of the variables of interest [Cameron and Trivedi, 2009]. Appying the "tetrad" approach proposed in Head et al. [2010] to Gamma (and Poisson) estimation as in table 3 partially addresses this problem, for it allows algebraic elimination of the importer-time and exporter-time effects (but not of the trading pair effects)[19].

The results of the regressions on the tetrads in table 3 are quite similar to those of the regressions based on the structural gravity equation. Again, the Gamma and the OLS regression support a positive and significant role of the immigrants in promoting the international trade of Spanish provinces while, as before, the PPML estimates do not support a significant role of either immigrants or emigrants. The magnitude of the elasticities, in this case, ranges between 0.14 and 0.187, a slightly wider range than in the previous case. As to the Gamma regression, the estimates based on the tetrads method are even closer to the Gamma regression in table 2 than are the OLS regressions: eliminating 575 indicator variables has not radically altered the estimates, suggesting the inconsistent estimates of the time-varying effects were not affecting $ln(Immi_{nit})$ and $ln(Emi_{nit})$; but it must be noted that the estimation of the time-varying effects counts on a much higher number of observations than the estimation of dyadic effects. Yet, it is not possible to definitely rule out the possibility that this occurs through the remaing time-invariant pair-specific effects.

[19] Efficient estimation using the tetrad method requires multi-way clustering at the level of the pair, of the importer-year and of the exporter-year, which I performed in the case of OLS using the code provided by Head et al. [2010] in their web appendix. Multi-way clustering is not available for Gamma and Poisson regression so I used clustering at the pair level for Gamma and robust standard errors for the Poisson estimation. I also compare the results of multi-way clustering and of clustering at the pair level in the case of OLS, showing that the significance in estimates is not radically affected. In any case, the interest here is more on the magnitude of the point estimates than on their the significance.

Table 3: Regression results - Tetrads

Model	Head et al. [2010]	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·187***	0·185***	−0·122	0·140***
	(0·053)	(0·052)	(0·103)	(0·031)
$ln(Emi_{nit})$	−0·0229	0·026	−0·127	0·038
	(0·052)	(0·040)	(0·110)	(0·027)
NID	0·244*	0·192*	0·179	0·194***
	(·114)	(0·100)	(0·175)	(0·066)
NED	0·016	0·042	−0·102	0·051
	(·069)	(0·065)	(0·138)	(0·044)
Trading pair effects	Yes°	Yes	Yes	Yes
Province-time effects	Yes°	No	No	No
Country-time effects	Yes°	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Constant	$-1{\cdot}02e{-}09$	31·411***		13·843***
	(·018)	(5·572)		(4·077)
N	14625	15195	15195	15195
r2		0·008		

Note: Reference importer is France, reference exporter is the province of Madrid. All dependent and independent variables are "tetraded" and demeaned with respect to the reference importer and exporter (see [Head and Mayer, 2014, Head et al., 2010]). Multi-way clustered standard errors in parentheses in column "Head et al. [2010]"; standard errors are clustered at the pair level in columns "OLS" and "GammaPML". Standard errors are robust in column "PPML".

°All variables tetraded and de-meaned by pair, which is equivalent to including the three sets of effects.

* p<0.1, ** p<0.05, *** p<0.01

While it is not possible to definitely exclude the possibility that the Gamma estimates are inconsistent, the similarity of the Gamma and OLS estimates seems to be a recurrent pattern in the different specifications reported in section 5.3, where the migration variables are articulated according to differing country groupings.

5.2 Tests on the underlying distribution of the errors

In this section I will address the underlying distribution of the errors through some tests based on Manning and Mullahy [2001] and Santos-Silva and Tenreyro [2006] to identify the most efficient model among the three. The coefficient of interest in the Park-type test suggested by Manning and Mullahy [2001] and Santos-Silva and Tenreyro [2006] and reported in table 4 is the λ in equation 9. Regressing the log of the squared residuals on the log of the fitted values of the OLS regression, the estimate for λ is 1.56. Because the MaMu test is in fact a modification of Park's test for heteroskedasticity, this result confirms that the OLS estimates are affected by some heteroskedasticity. Because the estimates of the MaMu test may be affected by heteroskedasticity as well as those of the gravity regressions[Santos-Silva and Tenreyro, 2006], **?** advise to rather rely on the PML estimates for λ. The coefficients for λ estimated by Poisson and Gamma PML are in both cases very close to 2. This result points at a constant coefficient of variation in the residuals; in this case the most efficient estimators are the homoskedastic OLS on logs, which is the maximum likelihood estimator in case it is available, and the Gamma PML [Manning and Mullahy, 2001]. This explains why, if heteroskedasticity is not too marked, the Gamma and OLS estimates are quite similar.

Table 4: Manning and Mullahy test on the underlying distribution of the errors

Model	OLS residuals	Poisson PML residuals	Gamma PML residuals
$ln(\hat{\mu})$	1·562∗∗∗	1·981∗∗∗	2·123∗∗∗
	(·008)	(0·130)	(0·006)
Constant	−0·486∗∗∗	15·728∗∗∗	−1·405∗∗∗
	(·0725)	(0·066)	(0·059)
N	15195	15195	15195
r2	0·702	0·014	0·922

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 5: Simple regression of the residuals on the covariates

Model	OLS residuals	Poisson PML residuals	Gamma PML residuals
$ln(Immi_{nit})$	0·485∗∗∗	0·312 ∗ ∗	0·354∗∗∗
	(0·115)	(0·126)	(0·095)
$ln(Emi_{nit})$	−0·190	−0·033	−0·038
	(0·119)	(0·091)	(0·076)
NID	0·636∗∗∗	0·367	0·465 ∗ ∗
	(0·222)	(0·233)	(0·186)
NED	−0·0322	0·081	0·065
	(0·160)	(0·142)	(0·119)
Constant	10·93 ∗ ∗∗	14·476∗∗∗	25·821∗∗∗
	(0·728)	(0·731)	(0·636)
N	15195	15195	15195
r2	0·0970	0·114	0·936

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

In table 5 I perform a similar exercise regressing the log of the squared residuals on the covariates. In the three models, the variance of the errors increases in the logs of the immigrant stocks. To detect whether this result is to be interpreted as a sign of functional

Table 6: RESET tests on the estimation methods

Model	OLS model	PPML model	GammaPML
Square of the fitted values	$-\cdot 241$	$-0{\cdot}1103$	$-1{\cdot}33e{-}13$
P-value	0·0004	0·1455	0·000
Cube of the fitted values	·0067	0·0107	$1{\cdot}42e{-}20$
P-value	0·0137	0·4352	0·000
Joint p-value	0·0000	0·0092	·

form mis-specification and non-constancy in the covariates, as suggested by case 3 in Head and Mayer [2014], in table **??** I report the p-values resulting from Ramsey [1969] RESET-tests on each estimation method, where, differently from [Santos-Silva and Tenreyro, 2006], I also include the cube of the fitted values. The results of the tests provide some evidence in favour of the hypothesis of functional form mis-specification and of non-linearity in the migration variables for Poisson and OLS. In the cases of OLS the p-values of the RESET tests are always below 5%. The tests for the Gamma PML, while very precise, result in coefficients for the square and cube of the fitted values that are so small that they actually confirm the consistency of the estimator. The results of the diagnostic test on the distribution of the errors, thus, coherently lead to selecting the Gamma PML as the consistent estimator. The Gamma estimator, with respect to the PPML, appears also more flexible in terms of distributional assumptions.

Table 7: Regression results - Non-linearity in migration

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	$0{\cdot}18*$	0·11	$0{\cdot}10*$
	(0·11)	(0·08)	(0·06)
$ln(Immi_{nit})^2$	$-0{\cdot}00$	$-0{\cdot}01$	0·01
	(0·02)	(0·01)	(0·01)
$ln(Emi_{nit})$	$-0{\cdot}03$	$-0{\cdot}04$	0·01
	(0·08)	(0·06)	(0·05)
$ln(Emi_{nit})^2$	0·00	0·01	$-0{\cdot}00$
	(0·01)	(0·01)	(0·01)
NID	0·19	$0{\cdot}27**$	$0{\cdot}19**$
	(0·13)	(0·12)	(0·07)
NED	0·02	0·03	0·05
	(0·08)	(0·07)	(0·05)
N	15195·00	15195·00	15195·00
r2	0·12		

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Based on the tests performed above, the question arises of whether the differences in the estimation results between the OLS, Poisson and Gamma estimates are due other forms of non-linearity or functional form mis-specification or simply to differences in the efficiency of the estimators.

The main problem with the Poisson PML is related with the possibility of conditional mean mis-specification implied by the Ramsey test in Table 6: indeed, while the distributional assumptions of independence and small probability may look peculiar for trade data, and the value of the constant in table 4 implies $h = 2.75$ in equation 9, i.e. overdispersion, these should not matter for the consistency of the estimator as long as the conditional mean is correctly specified(Wooldridge [2002]; see also Santos Silva and Tenreyro's "Log of gravity webpage"). The results of the Ramsey test in table 6 shed some doubts on this assumption.

As the residuals result roughly log-normally distributed, the homoskedastic OLS estimator would be the MLE estimator; according to the simulations performed by Head and Mayer [2014], it should also be more robust than the Poisson PML estimator to functional form mis-specification. Whether the heteroskedasticity that has been detected in the estimator has practical implications depends on the countervailing effects of the positive bias in the OLS estimator introduced by heteroskedasticity vis-à-vis the high number of effects included in the specification which drive the estimates towards zero. In the specifications carried out in this paper, the difference in the magnitude of the coefficients between the OLS and the Gamma PML is between 10 and 35%, so, assuming consistency of the Gamma estimator, this should be roughly the order of magnitude of the bias.

The reason for the persisting differences in the estimates of the OLS, Poisson and Gamma PML could be found in the non-linearity: including the squared terms of $ln(Immi_{nit})$ and $ln(Emi_{nit})$ (Table 7) in the model, however, has the effect of reducing the efficiency of the linear estimates, while the squared terms result both economically and statistically insignificant. The point estimates, while not statistically significant from zero, are more aligned. The results are similar when including a cubic term.

An alternative explanation for the divergence between the PPML and Gamma estimates could be found in the different weights attributed by the estimators to the trade flows of different sizes [Head and Mayer, 2014]: the Poisson estimator attributes greater weight to larger trade flows, those which are likely to rely less on immigrant and emigrant networks because they are mainly determined by geographic proximity, institutional similarity and common trade arrangements; by contrast, according to Santos-Silva and Tenreyro [2006]the Gamma estimator gives more weight to smaller (and noisier) trade flows, those which are more likely to rely on immigrant and emigrant networks. I could not find support to this explanation in a series of unreported regressions where I scaled the variables taking into account Eaton and Kortum [2002], where I split the immigrant and emigrant variables by their above-median and below median elements and where I restricted the analysis to the smallest 75% of the observations: in all cases, the results remained qualitatively unchanged [20]

Based on the considerations above, I will rely on the Gamma estimates to test hypotheses in the next section 5.3. For the sake of completeness and in order to show the systematic similarity of the OLS and Gamma estimates, however, in what follows I will report the results of the estimates for the three estimation methods.

5.3 Testing of the Hypotheses

In sections 2.5 and 4.3 I outlined a series of Hypotheses that can now be tested based on the identification of the OLS as the consistent and efficient model.

1.1 Immigrants' and emigrants' aggregate effects on trade

Hp 1.1 on the positive and significant effect of both immigrants and emigrants in promoting the trade of Spanish provinces finds only partial confirmation in the data. As regards the immigration side, the hypothesis is strongly confirmed. As shown in Table 2, the Gamma (and OLS) estimates of the coefficients of $ln(Immi_{nit})$ are positive and significant. According to this specification, increasing the immigrant population by 10% would increase the trade of Spanish provinces by 1.4%, a higher but comparable estimate than the one found by Peri and Requena-Silvente [2010] on an earlier time period[21]. Instead, the coefficient of $ln(Emi_{nit})$ is not statistically significantly different from zero. Thus, the hypothesis of a positive role of emigrants in promoting the trade of Spanish provinces proposed by Murat and Pistoiesi [2009] for the case of Italy does not find empirical support when looking at the local networks of emigrants. This does not exclude that the flows of information within the emigrants network be mainly determined at the national level (see below the tests on geographic proximity).

[20] An alternative but untestable explanation is that the differences are due to measurement error, which according to the estimates in Santos-Silva and Tenreyro [2006] would bias the Gamma estimates upward by 17%. In this case the PPML would be preferable as the bias would only be of 2%. Yet, even taking into account the possibility that measurement error would upward bias the Gamma estimates upward, the differences between the PPML and Gamma estimates persist; considering that the confidence intervals of the very inefficient PPML estimates in some cases do not include the OLS estimates, divergence in the estimates seem to be more likely to be associated with other forms of functional form mis-specification that should be further explored

[21] Similarly to Peri and Requena-Silvente [2010], the coefficient of the no-immigrant dummy (NID) does not result significantly different from zero in the OLS estimates. Instead, it remains positive and statistically significant in all specifications of the Poisson and the Gamma model, with comparable or even greater magnitudes than $ln(Immi_{nit})$. According to this result, the provinces with no immigrants would on average trade more than the provinces with at least one immigrant. NID assumes frequently the value of 1 by two main types of countries: those which enjoy particularly favourable fiscal conditions (e.g. Andorra, Luxembourg, Cyprus) and the very remote countries. Because of the smaller distances and closer institutional settings, the former group is likely to be more represented than the latter in Spanish trade data.

1.2/1.3 The role of cultural and institutional similarity

The two corollaries of Hp 1.1, that the immigrants' and emigrants' effect on the trade of Spanish provinces is stronger with more institutionally (Hp 1.2) and culturally distant (Hp. 1.3) countries are tested in Tables 8a and 8b.

In table 8a, only the immigrants from extra-EU countries result significant in increasing the trade of Spanish provinces in the Gamma and OLS estimates. Coherently with the findings in Table 2, no emigrant variable results statistically to determine the level of trade. This result is compatible with the idea that the role of immigrants as brokers of the flow of communication and as facilitators of the enforcement of contracts in international transactions is stronger with countries that do not share the same institutional setting, i.e. in this case are non-EU countries. The reason is that sharing the institutional setting or a regional trade agreement (RTA), as well as a common language, are factors that increase bilateral trade *per se*, independently from the immigrant population; the direct effect of these determinants is absorbed in the fixed effects.This is shown more explicitly in table 9, where the estimated pair-specific fixed effects from the OLS regression (more consistent than the Gamma estimates of the fixed effects) are regressed on a series of traditional gravity determinants [Cheng and Wall, 2005]. From this regression, Spain results to trade on average 20% more with EU countries and countries of the European Economic Area (EEA), and 63% more with OECD countries.

As regards language commonality, D_{lang} has a coefficient of 0.975: Spanish provinces trade about 97.5% more with Spanish-speaking countries, independently from the immigrants that they host from these countries. In addition to this, the Gamma estimates in table 8b shows that immigrants from Spanish-speaking countries increase trade with their origin countries to a much greater extent - 82% more - than do the immigrants from non-Spanish speaking countries. This strongly disconfirms hypothesis 1.3; notice that relying on the OLS estimates would have led to confirm the hypothesis, even if the magnitude of $ln(Immi_{nit}^{Spa})$ estimated by OLS is very similar to the Gamma estimate. This result could be explained by taking as a reference the random encounter model by Wagner et al. [2002](equation 7): sharing a common language may give the immigrant easier access to information on trading opportunities not just in the origin country but also in the host country and increase the capacity of the immigrant to successfully realize the trading opportunity, i.e. act as a factor that increases the probability p that an immigrant has the capacity to facilitate the exchange.

2.1 Sub-national heterogeneity in the multilateral resistance terms

Turning to the hypotheses about sub-national heterogeneity, Table 11 reports the results of the specifications that omit the province-time effects, but include province income, implying that the multilateral resistance terms do not vary by province. The point estimates of the immigration variables are clearly not very sensitive to this change, suggesting that the

Table 8: Regression results: cultural and institutional similarity

(a) Regression results Hp 1.2 - Institutional similarity: EU countries

Model	OLS	PPML	GammaPML
$ln(Immi_{nit}^{EU})$	0·123	−0·013	0·080
	(0·094)	(0·071)	(0·051)
$ln(Immi_{nit}^{NEU})$	0·173**	0·088	0·155***
	(0·071)	(0·064)	(0·038)
$ln(Emi_{nit}^{EU})$	0·001	−0·036	0·036
	(0·074)	(0·052)	(0·046)
$ln(Emi_{nit}^{NEU})$	−0·015	0·064*	−0·004
	(0·049)	(0·034)	(0·029)
NID	0·173	0·228**	0·211***
	(0·111)	(0·109)	(0·065)
NED	0·033	0·077	0·054
	(0·069)	(0·065)	(0·041)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

(b) Regression results Hp 1.3 - Cultural similarity

Model	OLS	PPML	GammaPML
$ln(Immi_{nit}^{Spa})$	0·234	−0·038	0·235**
	(0·216)	(0·235)	(0·109)
$ln(Immi_{nit}^{NSpa})$	0·156**	0·049	0·129***
	(0·064)	(0·049)	(0·033)
$ln(Emi_{nit}^{Spa})$	−0·002	0·074	−0·026
	(0·056)	(0·121)	(0·036)
$ln(Emi_{nit}^{NSpa})$	−0·015	0·029	0·017
	(0·055)	(0·032)	(0·033)
NID	0·173	0·221**	0·211***
	(0·113)	(0·108)	(0·065)
NED	0·027	0·082	0·054
	(0·071)	(0·063)	(0·043)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

Table 9: Determinants of the fixed effects (OLS estimates)

$ln(Y_i)$	1·292***
	(0·014)
$ln(Y_j)$	0·642***
	(0·011)
$ln(DIST)$	−1·051***
	(0·030)
D_{colo_tie}	−0·137
	(0·150)
D_{lang}	0·761***
	(0·148)
D_{common_border}	0·890***
	(0·065)
D_{EUEEA}	0·207***
	(0·046)
D_{OECD}	0·631***
	(0·043)
Constant	−25·598***
	(0·340)
N	15195
r2	0·516

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

Table 10: Joint significance of the province-time effects (Table 2)

Model	OLS	PPML	GammaPML
test	F(199, 3038) =1.90	chi2(196) = 1169.52	chi2(206) = 5.2e+05
p-value of the test	0.0000	0.0000	0.0000

bilateral immigration effect on trade is not strongly correlated with the average market access of the host province; this in a way confirms empirically that the immigration effect acts at the level of the bilateral costs of trade. The estimates for the emigration effect remain insignificant. As far as the whole of the gravity model is concerned, however, the province-time effects results jointly highly significantly different from zero with all estimation methods (table 10); while a likelihood ratio test is not reliable with clustered and robust standard errors, both the log-likelihood and Akaike information criterion statistics improve by a greater extent than the reduction in the degrees of freedom when province-time effects are added to the specifications. Thus, hypothesis 2.1 is fully supported by the data and the inclusion of the province-time effects is preferred on both theoretical and practical grounds, even if the practical effect on the variables of interest is minor.

Table 11: Regression results Hp 2.1 - Sub-national heterogeneity in the multilateral resistance terms

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·161***	0·0485	0·126***
	(0·0611)	(0·0535)	(0·0319)
$ln(Emi_{nit})$	−0·001	−0·005	0·004
	(0·0438)	(0·0329)	(0·0266)
$ln(Y_{it})$	0·384	0·660***	0·306
	(0·373)	(0·202)	(0·205)
NID	0·169	0·173	0·182***
	(0·110)	(0·115)	(0·0653)
NED	0·0275	0·0604	0·0355
	(0·0694)	(0·0686)	(0·0411)
N	15195	15195	15195
r2	0·077		
log-likelihood	−14886·3	−1·71e+07	−143904·7
AIC	30302·6	3·42e+07	294417·3
Trading pair effects	Yes	Yes	Yes
Province-time effects	No	No	No
Country-time effects	Yes	Yes	Yes

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

2.2 Geographic proximity

Table 12 reports the results of the regression which includes both the immigrants from a given country residing within the province and those residing outside the province. They confirm the hypothesis that it is local networks of immigrants, rather than more far-reaching networks, that determine the trade-facilitation effect by immigrants. The flows of information relevant to trade creation by emigrants, instead, are to be found at the level of nation-wide networks of expatriates; as in the specification about language commonality, the OLS identifies a similar magnitude but not a statistically significant effect. This suggests that the type of information relevant to the trade facilitation effect is quite different for immigrants and emigrants. Considering that the emigrants' effects is likely to be associated with a taste effect, this result implies that either the information or the tastes - a distinction between the two is unfortunately impossible with the data at stake - that drive trade are not province-specific but rather nation-wide. This may also bear implications with respect to the profile of the migrants able to exert the trade facilitation effect.

Table 12: Regression results Hp 2.2 - Geographic proximity

Model	OLS	PPML	GammaPML
$ln(Immi_{nit})$	0·154 ∗∗	0·045	0·144∗∗∗
	(0·064)	(0·048)	(0·032)
$ln(ImmiOut_{nit})$	−1·117	−0·544∗∗∗	−0·082
	(1·307)	(0·190)	(0·065)
$ln(Emi_{nit})$	−0·002	0·013	0·025
	(0·051)	(0·035)	(0·026)
$ln(EmiOut_{nit})$	0·524	−0·429	0·279∗∗∗
	(0·936)	(0·379)	(0·057)
NID	0·170	0·218 ∗∗	0·189∗∗∗
	(0·112)	(0·109)	(0·065)
NED	0·034	0·067	0·051
	(0·068)	(0·062)	(0·039)
Constant	14·490		10·635∗∗∗
	(14·577)		(0·956)
Trading pair effects	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes
N	15195	15195	15195
r2	0·119		

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

2.3 Sub-national heterogeneity in the immigrants' effects

Allowing the elasticity of immigration and emigration to vary by region (i.e. by *Comunidad Autonoma*, Hypothesis 2.3 about the heterogeneity in the immigrants and emigrant effects by region finds strong support. Table 13 shows that Immigrant and emigrant stocks result significant in only a few regions; according to the Gamma estimates, only in Castilla y Leon, Cantabria, Catalunia, Extremadura, Galicia, Islas Canarias, Navarra and Region de Murcia the immigrants are able to exert a positive effect on trade; the magnitude of the effect ranges from 0.137 to 0.32, with a peak of 0.742 in the Islas canarias. In one case, the Islas Baleares, the effect is negative but weakly significant. With the exception of the Gamma estimate for Asturias, the regions where the effect on trade is significant for immigrants do not find a symmetric effect for emigrants, which result significant only in one region. In table A.19 in the Appendix, I further address hypothesis 2.3 about subnational heterogeneity on the immigration and emigration effect on trade by performing province-specific estimates. Also in this case, and similarly to the findings by Bandyopadhyay et al. [2008] on ethnic groups, I get that the effect is very heterogeneous across provinces. The immigration effect results especially significant for the provinces of Almeria, Cordoba, Girona, Las Palmas, Salamanca and Segovia and its magnitude results much higher - between 0.27 and 0.80 - than the average estimate of 0.17. It is negative and weakly significant in the case of Bizkaya. This result usefully complements the aggregate information provided in previous specifications and confirms the importance of focusing on the province level, but should be taken carefully, considering the smaller number of observations available to estimate each province-specific immigrant effect (max ca. 500 per province).

3.1 Non-constancy in the migrants' effects

In tables 14a and 14b, the hypothesis of non-constancy of the elasticity is tested by running the estimates separately for groups of regions distinguished by their shares of immigrants and expatriates. In Table 14a provinces are grouped by their total share of immigrants (low, intermediate, high), and immigrants result having an effect on trade only by intermediate levels of foreign population residing.It is important to notice that the immigration and emigration shares are calculated on the whole of the migrating population, and not on a single country group. So high immigration provinces represent provinces where the overall population of foreigners is comparatively high in relation to the total population of the province; high expatriation provinces represent provinces which have lost a comparatively high portion of their resident population through emigration to any country in the world. The Gamma estimator shows a non-constant pattern in the migration effect when differentiated by foreign or expatriates population shares. The effect of immigrants on trade results stronger by lower shares of the immigrant population and decreases by increasing shares of foreign population, which may be interpreted as a sign of decreasing returns from immigration to trade, possibly induced by higher social costs associated with greater immigrant populations in a country which still has a relatively short experience in managing immigration.

Table 13: Regression results Hp 2.3 - Elasticities specific by *Comunidades Autonomas*

Model	OLS		PPML		Gamma PML	
	Immigrants	Emigrants	Immigrants	Emigrants	Immigrants	Emigrants
Aragón	−0·00711	−0·0158	−0·120	0·133	−0·0571	−0·00337
	(0·130)	(0·147)	(0·125)	(0·126)	(0·0808)	(0·102)
Cantabria	0·332	−0·0705	0·0810	0·0992	0·240**	−0·0946
	(0·233)	(0·169)	(0·0929)	(0·150)	(0·109)	(0·104)
Castilla y León	0·230**	−0·0436	0·0926	−0·259*	0·137**	0·0111
	(0·110)	(0·125)	(0·0798)	(0·139)	(0·0570)	(0·0813)
Castilla-La Mancha	0·0267	0·166	0·0973	0·0560	−0·00496	0·134*
	(0·0813)	(0·127)	(0·0759)	(0·139)	(0·0494)	(0·0774)
Cataluña	0·244**	−0·00782	−0·146*	0·0211	0·276***	−0·0292
	(0·0957)	(0·100)	(0·0753)	(0·0511)	(0·0645)	(0·0622)
Com. Valenciana	−0·0825	−0·0609	0·00851	−0·113*	−0·0998	−0·0501
	(0·104)	(0·0701)	(0·0799)	(0·0623)	(0·0674)	(0·0439)
Comunidad de Madrid	0·151	−0·0584	0·182**	0·162	0·137	−0·0393
	(0·156)	(0·168)	(0·0815)	(0·107)	(0·104)	(0·112)
Extremadura	0·265	−0·249	0·202	0·378**	0·320***	−0·257
	(0·205)	(0·289)	(0·141)	(0·177)	(0·124)	(0·162)
Galicia	0·293	−0·0272	0·0274	0·0561	0·268***	−0·0686
	(0·180)	(0·189)	(0·0930)	(0·0712)	(0·0852)	(0·103)
Illes Baleares	−0·347	0·0591	0·0283	−0·798	−0·444*	0·284
	(0·334)	(0·340)	(0·481)	(0·504)	(0·265)	(0·215)
Islas Canarias	0·638	0·269	0·792	0·122	0·742***	0·187
	(0·435)	(0·324)	(0·689)	(0·225)	(0·238)	(0·206)
La Rioja	0·109	−0·000220	0·130	0·0696	0·147	0·000867
	(0·334)	(0·0390)	(0·125)	(0·0743)	(0·193)	(0·0277)
Navarra	0·292	−0·0390	0·501***	−0·0586	0·204*	0·0360
	(0·196)	(0·223)	(0·184)	(0·139)	(0·121)	(0·147)
País Vasco	−0·103	−0·0243	−0·126	0·00161	−0·0927	0·00892
	(0·0844)	(0·0934)	(0·0862)	(0·0875)	(0·0606)	(0·0709)
Principado de Asturias	0·251	−0·127	−0·344*	0·766**	0·137	0·0398
	(0·275)	(0·195)	(0·182)	(0·353)	(0·146)	(0·116)
Región de Murcia	0·161	−0·00736	0·374***	−0·0270	0·222**	−0·0218
	(0·185)	(0·107)	(0·0893)	(0·0681)	(0·110)	(0·0675)
NID	0.143		0.162		0.151***	
	(0.110)		(0.106)		(0.0656)	
NED	0.0252		0.0307		0.0345	
	(0.0683)		(0.0541)		(0.0419)	
Constant	7.716***				13.87***	
	(0.249)				(0.216)	
N	15195		15195		15195	

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

Looking at table 14b, where the provinces are distinguished by their expatriation rates, the Gamma estimates, as well as the OLS, suggest that, by low emigration rates, immigrants do not play a statistically significant role in promoting trade. Instead, the expatriates from these provinces have a positive effect on trade, with a positive and significant coefficient of 0.168 in the GammaPML case. In the case of provinces with intermediate shares of emigrant population, the effect of the expatriates becomes negative and significant in both the cases of OLS and GammaPML. Immigrants in mid-expatriation provinces, instead, have a positive and significant role in promoting trade; the magnitude of their effect is 0.18 in the case of Gamma and 0.22 in the case of OLS. The effect of immigrants is even stronger in high-expatriation provinces: the coefficient is 0.38 in the case of Gamma and 0.44 in the case of OLS. In the same provinces, the effect of emigrants is insignificant in determining trade.

An explanation for this result which can't be definitely tested here is that the fact that provinces today result as high-expatriation provinces are actually mirroring longer-term dynamics of expatriation. The experience of emigration in the provinces may have created the social and institutional conditions that are conducive to the economic integration of a more diverse population and to the access to new markets abroad through enforcement and information effects, while the long experience in emigration has exhausted the opportunities available to emigrants, in particular those associated with the "taste effects". Notice that this interpretation is not substitutable with the hypothesis that has been put forward in the literature (es. **?**, that migration ancestry of a given ethnic group affects trade. Because the provinces are separated by the overall shares of emigrants, the mechanism is probably to be found at the level of the social sedimentation of the emigration experience. This explanation would also be compatible with the positive, but less sizeable role played by immigrants in mid-expatriation provinces and with the positive role played by emigrants in low expatriation provinces, i.e. where, according to this interpretation, the opportunities for entering new markets are not yet exhausted. The data do not allow distinguishing whether this goes through a taste effect, an enforcement or an information effect, but the first is likely to be important. The negative coefficient of the emigration variables in the mid expatriation provinces could be explained by the loss of human resources through emigration which also affects trading opportunities. The results in table 14b could also explain why, on the aggregate sample, the emigrant population does not result significant in affecting trade: the effects of the emigrants from low-expatriation provinces are offset by those of the emigrants in mid-expatriation provinces. Such an effect would logically imply a similarly negative effect of emigration in high-expatriation country. Indeed, the Gamma and OLS estimates are negative, while not significant.

Overall, the hypotheses of non-constancy in the immigrants' and emigrants' effect on trade in connection with external factors find strong support in the Gamma estimates. Comparing tables 14a and 14b with the estimates in table 7 above, however, it seems appropriate to conclude that the non-constancy is due to the interaction of the immigration and emigration variables with other factors associated with the characteristics of the location rather than to non-linearity in functional form of the variables themselves.

6 Conclusions

On the basis of the theoretical microfoundations of the gravity model, in this paper I identified the necessary ingredients for theory-consistent estimation of the gravity model: bilateral cost terms that are both pair-specific and time-varying (like immigration and emigration stocks); and effects at the importer-time, exporter-time and trading pair level. The practical importance of the inclusion of province-specific multilateral resistance terms (in this case the province-specific market access in each year) in predicting trade volumes has been confirmed for this specific dataset with an ad-hoc test.

The theory-consistent specification has been implemented through three estimation methods, whose consistency and efficiency for the data at stake have been compared on the basis of the tests proposed in **?** and Santos-Silva and Tenreyro [2006]. This procedure has led to identifying the Gamma PML as the most efficient and consistent estimator for the data at stake. Throughout the different specifications, the magnitudes of the OLS and Gamma PML estimates have resulted very close to each other, with the Gamma PML generally outperforming the OLS in terms of efficiency of the estimates and the OLS estimates exceeding those of the Gamma by between 10% and 35%. The Poisson PML has instead shown a different pattern of systematic non-significance in the estimates and in many cases estimates that were not comparable with the Gamma and OLS, to an extent which can hardly be solely explained with its lower efficiency for the data at stake, nor with the different weights attributed by the estimator to greater trade values. The main reason for the persisting differences in the estimates seems to be primarily found in functional form mis-specification which is likely to be associated with non-constancy in the immigration and emigration effects; in particular, non-constancy in the immigration and emigration effects has been identified across provinces with different overall immigration and expatriation shares.

Overall, the Gamma and OLS estimators robustly confirm a positive effect of immigrants on trade. The effect is stronger in provinces with low shares of the foreign population and it decreases by higher foreign population shares; it is stronger in the trade with more institutionally distant countries and with Spanish-speaking countries; it shows marked sub-national variation and appears to be due to localized rather than more far-reaching networks of immigrants. The networks of expatriates, instead, appear to affect trade through different mechanisms, with a strong role of nation-wide networks and negligible effects of the local networks; the effect of emigrants on trade is strong and positive in provinces with low expatriation levels, but it becomes negative in provinces with higher

Table 14: Regression results Hp 3.1 - Immigration and Emigration shares

(a) Shares of immigrants in provinces

Immigrants share	High			Intermediate			Low		
Model	OLS	GammaPML	PPML	OLS	GammaPML	PPML	OLS	GammaPML	PPML
$ln(Immi_{nit})$	0·025	0·032	0·169**	0·206**	0·110**	0·065	0·112	0·172**	0·001
	(0·110)	(0·061)	(0·078)	(0·093)	(0·045)	(0·071)	(0·112)	(0·067)	(0·115)
$ln(Emi_{nit})$	0·004	0·002	0·083**	0·057	0·089	−0·123*	−0·232*	−0·193**	0·029
	(0·051)	(0·029)	(0·039)	(0·093)	(0·055)	(0·063)	(0·124)	(0·080)	(0·064)
NID	−0·053	−0·047	0·112	0·216	0·197**	0·218*	0·193	0·277*	0·253
	(0·336)	(0·181)	(0·216)	(0·149)	(0·081)	(0·120)	(0·217)	(0·145)	(0·227)
NED	0·006	0·019	0·051	0·055	0·080	0·004	−0·061	−0·054	0·150
	(0·087)	(0·058)	(0·066)	(0·105)	(0·067)	(0·098)	(0·187)	(0·108)	(0·106)
Trading pair effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4460·000	4460·000	4460·000	6915·000	6915·000	6915·000	3505·000	3505·000	3505·000
r2	0·168			0·147			0·183		

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

(b) Shares of emigrants in provinces

Emigrants share	High			Intermediate			Low		
Model	OLS	GammaPML	PPML	OLS	GammaPML	PPML	OLS	GammaPML	PPML
$ln(Immi_{nit})$	0·444**	0·378***	0·044	0·220***	0·182***	0·114*	−0·067	−0·052	−0·144
	(0·192)	(0·099)	(0·145)	(0·082)	(0·046)	(0·068)	(0·096)	(0·057)	(0·102)
$ln(Emi_{nit})$	−0·130	−0·069	0·100*	−0·121**	−0·105***	−0·039	0·180*	0·168***	0·107
	(0·178)	(0·105)	(0·054)	(0·053)	(0·033)	(0·049)	(0·093)	(0·059)	(0·065)
NID	0·512*	0·527***	0·634***	0·077	0·152*	0·144	0·084	0·135	−0·025
	(0·286)	(0·166)	(0·217)	(0·145)	(0·089)	(0·168)	(0·195)	(0·125)	(0·197)
NED	−0·133	−0·049	−0·058	−0·151*	−0·122**	0·139	0·264**	0·281***	0·155**
	(0·222)	(0·133)	(0·135)	(0·091)	(0·058)	(0·109)	(0·114)	(0·071)	(0·079)
Trading pair effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-time effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-time effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2900·000	2900·000	2900·000	7000·000	7000·000	7000·000	4035·000	4035·000	4035·000
r2	0·190			0·163			0·176		

Standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

expatriation rates.

References

James E. Anderson. A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116, 1979.

James E. Anderson and Eric van Wincoop. Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93:170—-92, 2003.

Scott L. Baier and Jeffrey H. Bergstrand. Bonus vetus ols: A simple method for approximating international trade-cost e?ects using the gravity equation. *Journal of International Economics*, 77(1):77–85, 2009.

Richard Baldwin and Daria Taglioni. Gravity for dummies and dummies for gravity equations. *NBER working paper series*, 2007. Working Paper 12516, url-http://www.nber.org/papers/w12516.

Subhayu Bandyopadhyay, Cletus C. Coughlin, and Howard J. Wall. Ethnic networks and us exports. *Review of International Economics*, 16(1):199–213, 2008.

Ashok Deo Bardhan and Subhrajit Guhathakurta. Global links of subnational regions: Coastal exports and international networks. *Contemporary Economic Policy*, 22(225-236), 2004.

D. Boisso and M. Ferrantino. Economic distance, cultural distance, and openness in international trade: Empirical puzzles. *Journal of Economic Integration*, 12(4):456–484, 1997.

George Borjas. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118(4):1335–1374, 2003.

Hervé Boulhol and Clément Bosquet. Applying the glm variance assumption to overcome the scale- dependence of the negative binomial qgpml estimator, 2012.

Massimiliano Bratti, Luca De Benedictis, and Gianluca Santoni. On the pro–trade e?ects of immigrants, 2012.

Anthony Briant, Pierre-Philippe Combes, and Miren Lafourcade. Product complexity, quality of institutions and the pro-trade effect of immigrants, 2009.

A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics Using Stata*. Stata Press, College Station,TX, 2009.

Thomas Chaney. Distorted gravity – heterogeneous firms, market structure and the geography of international trade. *American Economic Review*, 98:1707—21, 2008.

I-Hui Cheng and Howard J. Wall. Controlling for heterogeneity in gravity models of trade and integration. *Federal Reserve Bank of St. Louis Review*, 87:49–63, 2005.

Catherine Co, Patricia Euzent, and Thomas Martin. The export effect of immigration into the usa. *Applied Economics*, 36:573–583, 2004.

Pierre-Philippe Combes, Miren Lafourcade, and Thierry Mayer. The trade-creating effects of business and social networks: evidence from france. *Journal of International Economics*, 66:1–29, 2005.

Hein de Haas. Migration transitions: a theoretical and empirical inquiry into the developmental drivers of international migration, 2010. IMI Working Paper No 24, International Migration Institute, University of Oxford.

James A. Dunlevy. The impact of corruption and language on the pro-trade effect of immigrants: evidence from the american states. *Review of Economics and Statistics*, 88 (1):182—-186, 2006.

James A. Dunlevy and William Hutchinson. The impact of immigration on american import trade in the late nineteenth and twentieth centuries. *Journal of Economic History*, 59: 1043–62, 1999.

Jonathan Eaton and Samuel Kortum. Technology, geography, and trade. *Econometrica*, 70(5):1741–1779, 2002.

Jonathan Eaton and Akiko Tamura. Bilateralism and regionalism in japanese and u.s. trade and direct foreign investment patterns. *Journal of the Japanese and International Economies*, 8:478–510, December 1994.

Robert C. Feenstra. *Advanced International Trade: Theory and Evidence.* Princeton University Press, Princeton, New Jersey, 2004.

S. Flisi and Marina Murat. The hub continent. immigrant networks, emigrant diasporas and fdi. *The Journal of Socio-Economics*, 40:796—805, 2011.

Sourafel Girma and Zhihao Yu. The link between immigration and trade: Evidence from the united kingdom. *Weltwirtschaftliches Archiv 138*, page 115–30, 2002.

David. M. Gould. Immigrant links to the home country: Empirical implications for u.s. bilateral trade flows. *The Review of Economics and Statistics*, 76(2):302–316, May 1994.

Zvi Griliches. Economic data issues. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, volume III. Elsevier Science Publishers BV, Amsterdam, 1986.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Cultural biases in economic exchange? *The Quarterly Journal of Economics*, pages 1095–1131, August 2009.

Keith Head and Thierry. Mayer. Gravity equations: Workhorse, toolkit, and cookbook. In Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, editors, *Handbook of International Economics*. Elsevier, 2014.

Keith Head, Thierry Mayer, and John Ries. The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14, 2010.

Elhanan Helpman, Marc Melitz, and Yona Rubinstein. Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, 123(2):441–487, 2008.

Mark G. Herander and Luz A. Saavedra. Exports and the structure of immigrant-based networks: The role of geographical proximity. *The Review of Economics and Statistics*, 87(2):323–335, May 2005.

David Hummels. Toward a geography of trade costs. Mimeo, Krannert School of Management, Purdue University, 1999.

Daniel Kaufmann, Aart Kraay, and Pablo Zoido-Lobaton. Governance matters. Policy Research Working Paper 2196, The World Bank, October 1999.

Willard Manning and John Mullahy. Estimating log models: to transform or not to transform? *Journal of health economics*, 20(4):461–494, 2001.

Marina Murat and Barbara Pistoiesi. Migrant networks: Empirical implications for the italian bilateral trade. *International Economic Journal*, 23(3):371–390., 2009.

Giovanni Peri and Francisco Requena-Silvente. The trade creation effect of immigrants: Evidence from the remarkable case of spain. *Canadian Journal of Economics*, 43(4), 2010.

Alejandro Portes and Julia Sensenbrenner. Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology*, 98:1320–50, 1993.

James L. Powell. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25:303—325, July 1984.

J. B. Ramsey. Tests for specification errors in classical linear least-squares analysis. *Journal of the Royal Statistical Association*, 71:350–371, 1969.

James E. Rauch. Networks versus markets in international trade. *Journal of International Economics*, 48(1):7–35, 1999.

James E. Rauch. Business and social networks in international trade. *Journal of Economic Literature*, 39(4):1177–1203, December 2001.

James E. Rauch and Vitor Trinidade. Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–30, 2002.

Stephen J. Redding and Anthony J. Venables. Economic geography and international inequality. *Journal of International Economics*, 62:53–82, 2004.

J.M.C. Santos-Silva and Silvana Tenreyro. The log of gravity. *Review of Economics and Statistics*, 88:641–58, 2006.

Bedassa Tadesse and Roger White. Do immigrants counter the effect of cultural distance on trade? evidence from us state-level exports. *Journal of Socio-Economics*, 37(6): 2304–2318, December 2008.

Jan Tinbergen. *Shaping the World Economy; Suggestions for an International Economic Policy.* Twentieth Century Fund, New York, 1962.

Don Wagner, Keith Head, and John Ries. Immigration and the trade of provinces. *Scottish Journal of Political Economy*, 49(5):507—525, 2002.

Jeffrey Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge (MA), 2002.

A Appendix

Table A.17: Total population, immigrants and emigrants by province (2010)

Province	Total pop.	Spanish	Foreigners	Share of immigrants over total popultation		% immigrant pop.	Emigrants	Share of emigrants over total population		% of emigrant pop.
TOTAL										
Spain	47021031	41273297	5747734	0.122	-	1.000	1408825	0.030	-	1.000
Alicante	1926285	1459186	467099	0.242	high	0.081	21371	0.011	low	0.015
Balears	1106049	863793	242256	0.219	high	0.042	14328	0.013	low	0.010
Almería	695560	544401	151159	0.217	high	0.026	27772	0.040	high	0.020
Girona	753046	590799	162247	0.215	high	0.028	9884	0.013	low	0.007
Tarragona	808420	658106	150314	0.186	high	0.026	10087	0.012	low	0.007
Castellón	604274	492009	112265	0.186	high	0.020	5267	0.009	low	0.004
Lleida	439768	359278	80490	0.183	high	0.014	11471	0.026	mid	0.008
Málaga	1609557	1334530	275027	0.171	high	0.048	33211	0.021	mid	0.024
Madrid	6458684	5378740	1079944	0.167	high	0.188	174819	0.027	mid	0.124
Murcia	1461979	1220114	241865	0.165	high	0.042	19607	0.013	mid	0.014
Guadalajara	251563	212359	39204	0.156	high	0.007	2247	0.009	low	0.002
S.C.Tenerife	1027914	874587	153327	0.149	high	0.027	72454	0.070	high	0.051
Barcelona	5511147	4705660	805487	0.146	high	0.140	104302	0.019	mid	0.074
LaRioja	322415	275735	46680	0.145	high	0.008	10237	0.032	mid	0.007
LasPalmas	1090605	936553	154052	0.141	mid	0.027	25548	0.023	mid	0.018
Zaragoza	973252	845610	127642	0.131	mid	0.022	15388	0.016	mid	0.011
Cuenca	217716	189747	27969	0.128	mid	0.005	2269	0.010	low	0.002

Continued...

Segovia	164268	143194	21074	0.128	mid	0.004	2304	0.014	mid	0.002
Valencia	2581147	2266752	314395	0.122	mid	0.055	36944	0.014	mid	0.026
Huesca	228566	200756	27810	0.122	mid	0.005	5063	0.022	mid	0.004
Teruel	145277	127643	17634	0.121	mid	0.003	3656	0.025	mid	0.003
Toledo	697959	613984	83975	0.120	mid	0.015	6627	0.009	low	0.005
Melilla	76034	67161	8873	0.117	mid	0.002	3527	0.046	high	0.003
Navarra	636924	565555	71369	0.112	mid	0.012	16766	0.026	mid	0.012
Soria	95258	85388	9870	0.104	mid	0.002	4421	0.046	high	0.003
Burgos	374826	340260	34566	0.092	mid	0.006	12122	0.032	mid	0.009
Araba/Álava	317352	289142	28210	0.089	mid	0.005	4139	0.013	low	0.003
C. Real	529453	483452	46001	0.087	mid	0.008	4175	0.008	low	0.003
Huelva	518081	475328	42753	0.083	mid	0.007	5200	0.010	low	0.004
Albacete	401682	369277	32405	0.081	mid	0.006	5129	0.013	low	0.004
Ávila	171896	159283	12613	0.073	mid	0.002	6005	0.035	mid	0.004
Granada	918072	853738	64334	0.070	mid	0.011	34317	0.037	high	0.024
Cantabria	592250	553049	39201	0.066	mid	0.007	25170	0.042	high	0.018
Valladolid	533640	500984	32656	0.061	mid	0.006	9005	0.017	mid	0.006
Gipuzkoa	707263	664814	42449	0.060	mid	0.007	19313	0.027	mid	0.014
Bizkaia	1153724	1085014	68710	0.060	mid	0.012	27011	0.023	mid	0.019
León	499284	473321	25963	0.052	mid	0.005	35339	0.071	high	0.025
Ourense	335219	318508	16711	0.050	mid	0.003	82134	0.245	high	0.058
Ceuta	80579	76584	3995	0.050	low	0.001	2132	0.026	mid	0.002
Salamanca	353619	336113	17506	0.050	low	0.003	23265	0.066	high	0.017
Asturias	1084341	1035055	49286	0.045	low	0.009	83041	0.077	high	0.059
Palencia	172510	165301	7209	0.042	low	0.001	5510	0.032	mid	0.004
Zamora	194214	186173	8041	0.041	low	0.001	14820	0.076	high	0.011
Pontevedra	962472	922678	39794	0.041	low	0.007	106279	0.110	high	0.075
Sevilla	1917097	1840007	77090	0.040	low	0.013	22326	0.012	low	0.016
Lugo	353504	339328	14176	0.040	low	0.002	50352	0.142	high	0.036
Cádiz	1236739	1188972	47767	0.039	low	0.008	19825	0.016	mid	0.014
Cáceres	415083	399767	15316	0.037	low	0.003	12705	0.031	mid	0.009
Badajoz	692137	668097	24040	0.035	low	0.004	8803	0.013	low	0.006
ACoruña	1146458	1107469	38989	0.034	low	0.007	128090	0.112	high	0.091
Córdoba	805108	779849	25259	0.031	low	0.004	13920	0.017	mid	0.010
Jaén	670761	650094	20667	0.031	low	0.004	9128	0.014	mid	0.006

Table A.19: Regression results Hp 2.2 - Province-specific elasticities

Province	OLS	OLS2	PPML	PPML2	GammaPML	GammaPML2
Immigrants						
Albacete	-0.197	-0.142	0.042	-0.174	-0.214**	-0.187
	(0.137)	(0.182)	(0.161)	(0.168)	(0.095)	(0.116)
Alicante	-0.117	-0.006	0.094	0.237	-0.109	-0.014
	(0.155)	(0.188)	(0.126)	(0.279)	(0.089)	(0.118)
Almeria	0.570**	0.504	0.479***	0.350**	0.494***	0.389*
	(0.268)	(0.393)	(0.095)	(0.146)	(0.157)	(0.218)
Asturias	0.254	0.118	-0.334*	-0.214	0.141	0.003
	(0.275)	(0.254)	(0.176)	(0.235)	(0.146)	(0.170)
Avila	-0.297	0.783	-0.430**	-0.316	-0.951**	0.279
	(0.728)	(0.878)	(0.183)	(0.392)	(0.423)	(0.623)
Badajoz	0.196	-0.397	0.284	-0.547*	0.306**	-0.319*
	(0.277)	(0.315)	(0.175)	(0.313)	(0.154)	(0.190)

Continued. . .

Islas Baleares	-0.341	0.212	0.047	0.852	-0.436	0.183
	(0.336)	(0.541)	(0.480)	(0.654)	(0.265)	(0.369)
Barcelona	-0.195	0.005	-0.209**	0.123	-0.223**	-0.025
	(0.149)	(0.167)	(0.092)	(0.114)	(0.090)	(0.105)
Burgos	0.167	0.222	0.228***	0.244*	0.149**	0.179*
	(0.106)	(0.165)	(0.088)	(0.126)	(0.068)	(0.092)
Caceres	0.338	0.180	0.217*	0.246	0.352*	0.248
	(0.304)	(0.310)	(0.117)	(0.197)	(0.193)	(0.217)
Cadiz	-0.414	0.153	-0.273*	-0.014	-0.352**	0.311
	(0.255)	(0.324)	(0.158)	(0.206)	(0.140)	(0.210)
Cantabria	0.345	0.333	0.099	0.161	0.248**	0.269*
	(0.233)	(0.277)	(0.092)	(0.117)	(0.109)	(0.150)
Castellon	0.038	0.180	0.050	0.283	-0.006	0.126
	(0.150)	(0.146)	(0.125)	(0.200)	(0.114)	(0.119)
Ciudad Real	0.181	0.121	0.291**	0.192*	0.171*	0.111
	(0.149)	(0.165)	(0.114)	(0.107)	(0.092)	(0.102)
Cordoba	0.279*	0.225	0.231	0.162	0.283***	0.235**
	(0.157)	(0.172)	(0.221)	(0.326)	(0.093)	(0.110)
A Coruna	0.182	0.177	0.185	0.087	0.192	0.274
	(0.168)	(0.270)	(0.126)	(0.179)	(0.118)	(0.175)
Cuenca	0.239	0.022	0.644**	-0.047	0.222*	-0.015
	(0.201)	(0.201)	(0.257)	(0.197)	(0.134)	(0.142)
Girona	0.519***	0.429**	0.163	0.336*	0.555***	0.463***
	(0.172)	(0.212)	(0.150)	(0.187)	(0.110)	(0.135)
Granada	0.159	0.079	-0.079	0.065	0.130	0.114
	(0.262)	(0.298)	(0.297)	(0.424)	(0.168)	(0.199)
Guadalajara	0.061	-0.178	0.133	0.333	0.012	-0.147
	(0.150)	(0.190)	(0.153)	(0.228)	(0.092)	(0.120)
Guipuzcoa	-0.016	-0.092	-0.107	0.014	-0.034	-0.127
	(0.155)	(0.161)	(0.132)	(0.159)	(0.109)	(0.128)
Huelva	0.042	0.122	0.349*	-0.042	0.014	-0.005
	(0.257)	(0.409)	(0.186)	(0.194)	(0.157)	(0.263)
Huesca	0.036	0.087	0.536*	0.518	-0.001	0.030
	(0.210)	(0.227)	(0.291)	(0.407)	(0.144)	(0.158)
Jaen	0.277	0.578***	0.033	0.184	0.254*	0.587***
	(0.221)	(0.221)	(0.278)	(0.448)	(0.130)	(0.131)
La Rioja	0.128	-0.002	0.154	-0.041	0.170	-0.007
	(0.336)	(0.360)	(0.125)	(0.111)	(0.195)	(0.202)
Leon	0.121	0.113	0.612**	0.650**	0.096	0.223
	(0.283)	(0.317)	(0.280)	(0.279)	(0.197)	(0.223)
Lleida	0.224	-0.035	0.264**	-0.071	0.188**	-0.070
	(0.147)	(0.178)	(0.110)	(0.154)	(0.094)	(0.116)
Lugo	0.297	0.226	0.023	-0.576**	0.360**	0.253
	(0.249)	(0.307)	(0.162)	(0.244)	(0.161)	(0.185)
Madrid	0.161	0.398	0.207**	0.058	0.140	0.425**
	(0.159)	(0.261)	(0.083)	(0.178)	(0.105)	(0.173)
Malaga	-0.033	-0.136	0.389	-0.077	-0.090	-0.139
	(0.240)	(0.329)	(0.297)	(0.305)	(0.177)	(0.237)
Murcia	0.166	0.224	0.395***	0.432***	0.225**	0.315**
	(0.187)	(0.218)	(0.091)	(0.124)	(0.110)	(0.138)
Navarra	0.309	0.312	0.523***	0.341*	0.216*	0.162
	(0.196)	(0.203)	(0.184)	(0.197)	(0.121)	(0.128)
Ourense	0.582	0.458	0.415	0.858**	0.446**	0.291
	(0.460)	(0.572)	(0.363)	(0.412)	(0.192)	(0.224)
Palencia	0.208	0.243	0.326**	0.001	0.143	0.210*
	(0.218)	(0.251)	(0.157)	(0.097)	(0.114)	(0.121)

Continued. . .

Las Palmas	0.802**	0.125	1.392*	1.915	0.912***	0.351
	(0.385)	(0.452)	(0.723)	(1.331)	(0.293)	(0.361)
Pontevedra	-0.021	-0.018	-0.112	0.071	-0.030	-0.048
	(0.227)	(0.268)	(0.117)	(0.235)	(0.169)	(0.212)
Salamanca	0.635*	0.389	0.579**	-0.340	0.583***	0.423*
	(0.367)	(0.404)	(0.230)	(0.287)	(0.192)	(0.221)
S.C. Tenerife	0.524	-0.385	-0.973	1.788	0.581	0.166
	(0.814)	(0.919)	(1.546)	(1.715)	(0.385)	(0.430)
Segovia	0.677**	0.789**	0.376***	0.289*	0.494***	0.614***
	(0.283)	(0.314)	(0.090)	(0.162)	(0.133)	(0.160)
Sevilla	-0.079	-0.447*	0.170	-0.412**	-0.088	-0.466***
	(0.200)	(0.237)	(0.178)	(0.179)	(0.128)	(0.147)
Soria	0.212	0.651**	-0.049	0.179	0.162	0.479***
	(0.302)	(0.291)	(0.161)	(0.244)	(0.137)	(0.143)
Tarragona	0.284	0.142	0.130	0.297	0.406***	0.317
	(0.182)	(0.264)	(0.102)	(0.184)	(0.133)	(0.204)
Teruel	0.031	0.076	-0.101	-0.206	-0.057	-0.102
	(0.218)	(0.239)	(0.105)	(0.153)	(0.133)	(0.139)
Toledo	-0.157	0.036	-0.203	0.089	-0.186*	-0.014
	(0.185)	(0.186)	(0.128)	(0.224)	(0.111)	(0.130)
Valencia	-0.222	-0.042	0.040	0.333**	-0.207**	-0.027
	(0.159)	(0.171)	(0.103)	(0.163)	(0.094)	(0.122)
Valladolid	0.245	0.319	-0.163	-0.062	0.137	0.228
	(0.231)	(0.322)	(0.199)	(0.289)	(0.125)	(0.163)
Vizcaya	-0.272*	-0.140	-0.147	-0.050	-0.226*	-0.119
	(0.161)	(0.145)	(0.149)	(0.174)	(0.123)	(0.123)
Zamora	-0.106	-0.171	0.122	-0.438**	-0.120	-0.179
	(0.291)	(0.330)	(0.172)	(0.186)	(0.188)	(0.211)
Zaragoza	-0.026	-0.219	-0.171	-0.305**	-0.019	-0.186
	(0.195)	(0.248)	(0.110)	(0.138)	(0.117)	(0.149)
Emigrants						
Albacete	0.146	0.204	0.256	0.273	0.153	0.212*
	(0.202)	(0.209)	(0.217)	(0.233)	(0.123)	(0.124)
Alicante	0.091	0.140*	0.006	0.044	0.109**	0.153***
	(0.064)	(0.072)	(0.077)	(0.068)	(0.047)	(0.054)
Almeria	0.680***	0.670***	0.509***	0.488***	0.614***	0.599***
	(0.263)	(0.237)	(0.133)	(0.134)	(0.163)	(0.159)
Asturias	-0.112	-0.161	0.771**	0.714**	0.052	-0.022
	(0.193)	(0.207)	(0.352)	(0.343)	(0.115)	(0.131)
Avila	-0.347	-0.346	-0.162	-0.124	-0.200	-0.198
	(0.583)	(0.594)	(0.192)	(0.232)	(0.409)	(0.416)
Badajoz	-0.352	-0.741	0.615***	0.387**	-0.323	-0.684***
	(0.472)	(0.469)	(0.184)	(0.190)	(0.240)	(0.244)
Islas Baleares	0.090	0.295	-0.807	-0.539	0.309	0.495**
	(0.342)	(0.317)	(0.503)	(0.463)	(0.216)	(0.221)
Barcelona	-0.013	0.083	0.019	0.109*	0.012	0.112*
	(0.101)	(0.102)	(0.060)	(0.056)	(0.064)	(0.068)
Burgos	0.124	0.146	0.000	0.006	0.155	0.165
	(0.180)	(0.185)	(0.119)	(0.114)	(0.119)	(0.119)
Caceres	-0.081	-0.240	-0.131	-0.224	-0.136	-0.223
	(0.325)	(0.342)	(0.173)	(0.212)	(0.213)	(0.227)
Cadiz	-0.473**	-0.272	-0.439**	-0.478**	-0.583***	-0.321*
	(0.236)	(0.217)	(0.185)	(0.188)	(0.170)	(0.170)
Cantabria	-0.044	-0.050	0.105	0.077	-0.077	-0.065
	(0.169)	(0.172)	(0.150)	(0.150)	(0.104)	(0.103)

Continued...

Castellon	-0.031	0.119	-0.045	0.040	-0.065	0.091
	(0.103)	(0.112)	(0.075)	(0.084)	(0.066)	(0.076)
Ciudad Real	0.071	-0.005	-0.058	-0.040	0.051	-0.010
	(0.221)	(0.229)	(0.261)	(0.313)	(0.139)	(0.144)
Cordoba	0.139	0.045	-0.052	-0.077	0.083	-0.007
	(0.201)	(0.239)	(0.206)	(0.228)	(0.132)	(0.147)
A Coruna	0.087	0.064	-0.166	-0.210	0.092	0.123
	(0.261)	(0.255)	(0.259)	(0.274)	(0.183)	(0.187)
Cuenca	0.284	0.039	0.269	0.191	0.354*	0.122
	(0.365)	(0.349)	(0.282)	(0.251)	(0.214)	(0.205)
Girona	-0.060	-0.075	0.030	0.056	-0.092	-0.100
	(0.163)	(0.162)	(0.109)	(0.124)	(0.113)	(0.117)
Granada	0.196	0.212	0.008	0.149	0.162	0.240*
	(0.193)	(0.204)	(0.259)	(0.310)	(0.139)	(0.141)
Guadalajara	0.214	0.103	-0.203	-0.159	-0.085	-0.182
	(0.302)	(0.305)	(0.250)	(0.286)	(0.182)	(0.181)
Guipuzcoa	0.009	0.022	0.052	0.222	0.037	0.065
	(0.174)	(0.216)	(0.209)	(0.193)	(0.128)	(0.134)
Huelva	0.028	0.088	0.017	-0.459*	0.129	0.070
	(0.479)	(0.471)	(0.339)	(0.263)	(0.303)	(0.339)
Huesca	0.118	0.101	-0.405*	-0.514*	0.069	0.051
	(0.254)	(0.261)	(0.245)	(0.288)	(0.182)	(0.187)
Jaen	-0.280	-0.016	0.210	0.197	-0.172	0.144
	(0.436)	(0.386)	(0.407)	(0.415)	(0.228)	(0.201)
La Rioja	0.005	-0.017	0.074	0.024	0.005	-0.020
	(0.039)	(0.045)	(0.074)	(0.061)	(0.028)	(0.031)
Leon	0.558**	0.565**	0.559	0.493	0.837***	0.852***
	(0.270)	(0.268)	(0.470)	(0.509)	(0.234)	(0.217)
Lleida	0.351*	0.222	0.156	-0.004	0.337***	0.219*
	(0.206)	(0.205)	(0.182)	(0.171)	(0.130)	(0.130)
Lugo	-0.200	-0.185	0.353	0.336	-0.406	-0.368
	(0.453)	(0.430)	(0.422)	(0.402)	(0.261)	(0.259)
Madrid	-0.045	0.054	0.154	0.094	-0.024	0.100
	(0.166)	(0.195)	(0.108)	(0.106)	(0.110)	(0.119)
Malaga	-0.035	-0.068	-0.087	-0.363**	-0.020	-0.032
	(0.167)	(0.193)	(0.214)	(0.177)	(0.122)	(0.134)
Murcia	0.014	0.040	-0.024	-0.031	-0.007	0.034
	(0.108)	(0.117)	(0.068)	(0.074)	(0.067)	(0.071)
Navarra	-0.013	-0.038	-0.050	-0.200	0.060	-0.002
	(0.224)	(0.254)	(0.137)	(0.146)	(0.148)	(0.169)
Ourense	0.364	0.175	-0.335	-0.348	0.353	0.185
	(0.543)	(0.505)	(0.534)	(0.457)	(0.301)	(0.298)
Palencia	-0.356	-0.436	-0.019	-0.139	-0.267	-0.360
	(0.362)	(0.362)	(0.195)	(0.201)	(0.232)	(0.233)
Las Palmas	0.084	-0.266	0.242	0.390	-0.020	-0.315
	(0.380)	(0.393)	(0.454)	(0.508)	(0.272)	(0.331)
Pontevedra	-0.138	-0.172	0.111*	0.119	-0.090	-0.146
	(0.184)	(0.178)	(0.063)	(0.081)	(0.132)	(0.140)
Salamanca	0.028	-0.122	0.761*	0.712	0.013	-0.026
	(0.330)	(0.340)	(0.416)	(0.441)	(0.261)	(0.262)
S. C. Tenerife	0.438	-0.031	-0.055	0.555	0.386	0.066
	(0.478)	(0.485)	(0.216)	(0.557)	(0.292)	(0.356)
Segovia	-0.257	-0.266	-0.121	-0.124	-0.258	-0.326
	(0.490)	(0.492)	(0.236)	(0.239)	(0.226)	(0.225)
Sevilla	0.097	-0.051	0.366	0.174	0.048	-0.096
	(0.274)	(0.263)	(0.312)	(0.209)	(0.166)	(0.160)

Continued...

Soria	0.186	0.628*	0.232	0.539	0.281	0.677***
	(0.356)	(0.329)	(0.318)	(0.409)	(0.228)	(0.226)
Tarragona	-0.042	-0.112	0.105	0.139	-0.094	-0.127
	(0.199)	(0.192)	(0.179)	(0.176)	(0.124)	(0.125)
Teruel	0.025	0.012	0.078	0.059	0.113	0.094
	(0.408)	(0.364)	(0.567)	(0.584)	(0.293)	(0.287)
Toledo	0.374	0.416	0.064	0.109	0.427***	0.473***
	(0.271)	(0.280)	(0.311)	(0.267)	(0.161)	(0.160)
Valencia	-0.251	-0.128	-0.348**	-0.373**	-0.213**	-0.093
	(0.179)	(0.192)	(0.151)	(0.151)	(0.099)	(0.109)
Valladolid	-0.155	-0.096	-0.321***	-0.337***	-0.189	-0.139
	(0.176)	(0.236)	(0.106)	(0.106)	(0.124)	(0.137)
Vizcaya	0.135	0.210	-0.095	-0.088	0.141	0.195
	(0.180)	(0.194)	(0.161)	(0.165)	(0.138)	(0.149)
Zamora	-0.606	-0.722	0.902	0.793	-0.239	-0.356
	(0.465)	(0.444)	(1.045)	(0.857)	(0.384)	(0.395)
Zaragoza	-0.095	-0.181	0.200	0.173	-0.090	-0.146
	(0.117)	(0.135)	(0.125)	(0.121)	(0.097)	(0.105)
NID	0.191*	0.219*	0.284***	0.218**	0.201***	0.221***
	(0.114)	(0.116)	(0.097)	(0.106)	(0.068)	(0.069)
NED	0.078	0.070	0.041	0.030	0.073*	0.078*
	(0.072)	(0.071)	(0.057)	(0.056)	(0.044)	(0.043)
Constant	7.622***	7.842***			13.665***	13.693***
	(0.309)	(0.333)			(0.071)	(0.060)
Country-time effects	Yes	Yes	Yes	Yes	Yes	Yes
Province-time effects	No	Yes	No	Yes	No	Yes
Pair effects	Yes	Yes	Yes	Yes	Yes	Yes
N	15195	15195	15195	15195	15195	15195
r2	0.091	0.130				

Table A.15: List of Countries

Europe	Americas	Asia	Africa	Transition economies	Oceania
Andorra	Argentina	Bangladesh	Algeria	Bulgaria	Australia
Austria	Brazil	China	Angola	Croatia	
Belgium	Chile	India	Dem. Rep. of Congo	Czech Republic	
Cyprus	Dominican Rep.	Iran	Egypt	Hungary	
Denmark	Canada	Israel	Guinea	Lithuania	
Finland	Colombia	Lebanon	Morocco	Poland	
France	Cuba	Japan	Nigeria	Romania	
Germany	Mexico	Pakistan	Syria	Russia	
Greece	Panama	Philippines	South Africa	Slovakia	
Ireland	Peru	South Korea	Tunisia	Slovenia	
Italy	USA			Ukraine	
Luxembourg	Venezuela				
Malta					
The Netherlands					
Norway					
Portugal					
Sweden					
Switzerland					
Turkey					
UK					

Table A.16: Data Sources

Variable	Description	Units	Years	Source	URL
X_{nit}	Exports	Thousands of euros	1995–2011	Datacomex	http://datacomex.comercio.es/principal_comex_es.aspx. Full database at 3-digit disaggregation requested and received by email
$Immi_{nit}$	Stock of residents per province originating from a certain country	Persons	1996, 1998–2011	INE – "Población extranjera por sexo, comunidades y provincias y nacionalidad"	http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe245&file=inebase&L=0
Emi_{nit}	Stock of natives of a certain province residing in a certain country	persons	2006-2011	Censo Electoral de españoles residentes en el extranjero (CERA) por provincia de inscripción y país de residencia	http://www.ine.es/ss/Satellite?c=Page&cid=1254735793323&pagename=CensoElectoral%2FINELayout&L=0
Y_{nt}	Country GDP at current prices	Billions of US dollars	1995–2011	IMF World Economic Outlook Database	http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/index.aspx
Y_{it}	Province output at current prices	Thousands of euros	1995-2010	PIB a precios de Mercado precios Corrientes	http://www.ine.es/jaxi/menu.do?L=0&type=pcaxis&path=%2Ft35%2Fp010&file=inebase
$DIST$	Great circle distance between Madrid and the partner country capital city	km			http://www.chemical-ecology.net/java/lat-long.htm

Table A.18: Top origin and destination countries in 2010

A.Immigration: Top 15 origin countries, 2010

Country	Immigrants	Gender ratio	Age ratio	Education ratio	Employment ratio	Duration-of-stay ratio
1.Romania	831235	.6897932	3.179868	.6057692	1.987028	3.252971
2.Morocco	754080	.620505	2.894016	.2727942	1.183971	.6356112
3.Ecuador	399586	1.053598	2.514053	.538374	2.236376	3.031104
4.UK	387677	1.042696	2.351648	.9857143	.5566188	.4882227
5.Colombia	292641	1.478411	3.311899	.8573672	1.248511	2.234731
6.Bolivia	213169	1.15331	3.065789	1.063973	1.288889	1.586498
7. Germany	195824	1.083444	3.497498	1.148885	.7541121	.2828179
8. Italy	184277	.5826446	3.807531	1.371849	.9947917	.5593923
9. Bulgaria	169552	.7585185	3.844898	.6437588	1.75406	2.184783
10. China	158244	.803681	3.083333	.4401973	1.654628	.7176471
11. Portugal	142520	1.068939	3.193093	.1971067	.8730716	.1649802
12. Peru	140182	1.4389	4.543982	1.68623	1.424089	.7410388
13. Argentina	132249	1.054078	2.606226	1.480842	.9503367	.4862593
14. Francia	123870	1.168254	3.368023	.9148555	1.09919	.1192647
15. Brazil	117808	1.859375	3.318584	.9729364	.9972715	.5711207

B.Emigration: Top 15 destination countries, 2010

Country	Immigrants	Gender ratio	Age ratio	Education ratio	Employment ratio	Duration-of-stay ratio
1. Argentinaf	309523	1.159566122	0.991431207	-	-	-
2. France	164854	1.231658	1.273929	.4695855	.5418145	.0352967
3. Venezuelaf	139830	0.999325454	1.229308346	-	-	-
4. Germany	85700	.6560102	2.6108f	.4018945	-	-
5. Brazilf	80285	0.938242526	1.208772085	-	-	-
6. Switzerland	73277	.9812543	5.052161	.5997626	2.558361	.087425
7. Mexico	70173	.8634261	1.389786	2.071969	.9472904	-
8. Cubaf	70106	1.301930085	0.455867289	-	-	-
9. USA	57678	1.277718	2.638894	2.088229	.9939255	.2016603
10. UK	50727	1.477968	2.502992	2.316668	1.756874	-
11. Uruguayf	49422	1.160277815	1.057324571	-	-	-
12. Belgium	36222	1.141806	2.681823	.2854206	.6084685	.1230672
13. Chilef	33056	1.057610844	1.165552088	-	-	-
14. Andorraf	21319	1.019035178	3.021933086	-	-	-
15. Netherlands	15262	1.087178	3.344953613f	1.369096	-	-

Age ratio $= \frac{\text{Persons aged 25-64}}{\text{(Persons aged 15-24)+(Persons aged 65+)}}$; Education ratio $= \frac{\text{(ISCED codes 3 to 6}}{\text{(ISCED codes 0-2)}}$; Employment ratio $= \frac{\text{(Employed)}}{\text{(Unemployed)+(Inactive)}}$; Duration-of-stay ratio $= \frac{\text{stays}\leq\text{5years}}{\text{stays}>\text{5years}}$. Source: Own elaboration on INE and OECD-DIOC data.